TRANSPORTATION RESEARCH
RECORD

No. 1306

Planning and Administration

In-Vehicle Information Systems: Modeling Traffic Networks and Behavioral Considerations 1991

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1991

Transportation Research Record 1306
Price: $17.00

Subscriber Category
IA planning and administration

TRB Publications Staff
Director of Publications: Nancy A. Ackerman
Senior Editor: Naomi C. Kassabian
Associate Editor: Alison G. Tobias
Assistant Editors: Luanne Crayton, Kathleen Solomon, Norman Solomon
Graphics Coordinator: Diane L. Ross
Production Coordinator: Karen S. Waugh
Office Manager: Phyllis D. Barber
Production Assistant: Betty L. Hawkins

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

In-vehicle information systems : modeling traffic networks and behavioral considerations.
p. cm.—(Transportation research record ; no. 1306)
"A peer-reviewed publication of the Transportation Research Board."
ISBN 0-309-05114-2
1. Automobiles—Instruments—Display systems. 2. Traffic flow—Computer simulation. 3. Traffic engineering—Computer simulation. I. National Research Council
(U.S.). Transportation Research Board. II. Series:
Transportation research record ; 1306.
TE7.H5 no. 1306
[TL272.5]
388 s—dc20
[629.27'3] 91-33887
CIP

Sponsorship of Transportation Research Record 1306

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Chairman: Sally Hill Cooper, Virginia Department of Transportation

Transportation Economics and Data Forecasting Section
Chairman: Edward Weiner, U.S. Department of Transportation

Committee on Traveler Behavior and Values
Chairman: Ryuichi Kitamura, University of California-Davis
Elizabeth Deakin, Bernard Gerardin, Thomas F. Golob, David T. Hartgen, Joel L. Horowitz, Lidia P.Kostyniuk, Martin E. H. Lee-Gosselin, Hani S. Mahmassani, Patricia L. Mokhtarian, Frank Southworth, Peter R. Stopher, Janusz Supernak, Antti Talvitie, Mary Lynn Tischer, Carina Van Knippenberg, Martin Wachs, William Young

Committee on Transportation Supply Analysis
Chairman: Hani S. Mahmassani, University of Texas at Austin
David E. Boyce, Yupo Chan, Carlos F. Daganzo, Mark S. Daskin, Randolph W. Hall, William C. Jordan, Eric J. Miller, Anna Nagurney, Jossef Perl, Earl R. Ruiter, K. Nabil A. Safwat, Yosef Sheffi, Mark A. Turnquist

GROUP 5—INTERGROUP RESOURCES AND ISSUES
Chairman: William J. Harris, Jr., Texas A&M University

Task Force on Advanced Vehicle and Highway Technologies
Chairman: Daniel Brand, Charles River Associates Inc.
David E. Boyce, Richard P. Braun, Peter Davies, Shaun S. Devlin, Robert D. Ervin, Michael M. Finkelstein, Robert L. French, Franz K. Gimmler, Thomas A. Griebel, William J. Harris, Jr., Jack L. Kay, Bill M. McCall, Lawrence G. O'Connell, Robert E. Parsons, Stephen Edwin Rowe, Lyle Saxton, Joseph L. Schofer, William M. Spreitzer, Burton W. Stephens, Joseph M. Sussman, John Vostrez, David K. Willis

James A. Scott, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1990.

Transportation Research Record 1306

Contents

Foreword

Cascetta et al. discuss a dynamic assignment model for a general network, which follows a nonequilibrium approach in which flow fluctuations are modeled as a stochastic process and includes a model of dynamic network loading for computing within-day variable area flows from path flows. The sensitivity and operational characteristics of the model are tested by analyzing some effects of control measures on a small realistic network. The authors state the results show the proposed model is a valid tool to estimate the effectiveness of some traffic engineering measures and/or information systems.

In the second paper, Hamerslag and van Berkum discuss the use of road transport informatics, which aims to optimize the utilization of existing facilities in the transportation system serving three main goals: alleviation of congestion, diminution of air pollution, and reduction in incidents. The authors discuss providing the road user with information whereby stochastic and deterministic assignments are compared for both networks with and without congestion. To let information have also effect on destination choice and the spatial distribution of activities, the assignment models were combined with different distribution models. Simulations show that the amount of kilometers driven decreases when travelers are provided with better and more information.

In the third paper, van Vuren and Walting discuss the application of the concept of multiple user class equilibrium assignment to the modeling of route guidance systems. In particular its role in modeling guided and unguided drivers is discussed, as well as its ability to lead to guidance strategies that are effective even with a high proportion of drivers equipped.

In the fourth paper, Smith and Ghali describe a dynamic assignment/control model for evaluating urban traffic control schemes and route guidance strategies. The model is based on the dynamic assignment program CONTRAM for each fixed-road network subject to a given overall pattern of dynamic demand. The model evaluates all eight combinations of four responsive traffic control policies and two route guidance strategies on the assumption that the route guidance is accurately obeyed. The model given results for all required demand.

In the next paper, van Grol and Bakker discuss the linear processor array (LPA), which is a one-dimensional parallel computer with highspeed buffered interconnections between each pair of neighboring process boards, parallel accessible by both a control board and a general host computer, forms a transparent concept for the programmer. The optimally configured boards together with the high speed intercommunication allow a cost/performance improvement of a factor of 100 compared with a minisupercomputer like the Convex C1.

In the sixth paper, Papageorgiou and Messmer discuss a deterministic, microscopic modeling framework for dynamic traffic phenomena or networks consisting of freeways and/or urban streets for nonelastic but time varying traffic demands. A feedback methodology is applied to the network model so as to establish dynamic traffic assignment conditions. More specifically, a multivariable feedback regulation with integral parts and simple controller are developed and tested for a particular network traffic model.

In the paper, "In-Vehicle Information Systems: Modeling Traffic Networks, and Behavioral Considerations", Bonsall and Perry argue that because that sources of data in drivers reaction to route guidance is sketchy an interactive route choice simulator might provide substitute data. The use of IGOR to collect data under the European Drive initiative is described. The authors state that the acceptance of an item of advice depends on its (objective) quality, time drivers knowledge of network, and on the extent to which advice is corroborated by other evidence. The value of IGOR and its results are discussed.

In the paper, "Comparative Assessment of Origin-Based and En Route Real-time Information Under Alternative User Behavior Rules", Mahmassani and Shen-te Chen discuss the effects of real-time traffic information, supplied at the origin of the trip or along the way (enroute) on the system's performance under alternative behavior rules governing path selection in the network. Simulation experiments are performed to investigate the effect on overall performance as well as the incidence of benefits (costs) across user information groups of four experimental factors, behavioral rules, governing users' response to real-time information, sources of information, consisting of point-of-departure or in-vehicle (or both), prevailing "initial" conditions in the system and market penetration, i.e., the fraction of users with access to real-time information in the network. The results of these simulation experiments are discussed.

In the last paper, "Laboratory Assessment of Driver Route Diversion In Response to In-Vehicle Navigation and Motorist Information Systems", Allen et al. describe the first phase of a two part study of driver use of in-vehicle navigation systems. The second phase will apply driver behavior data to a traffic simulation model. The objective of the driver behavior experiment was to compare the effect of four navigation systems on driver diversion decision when faced with traffic congestion. Three of the systems were based on a heading up map display. The fourth system consisted of simplified symbolic directions and distance to change information.

Evaluation of Control Strategies Through a Doubly Dynamic Assignment Model

E. CASCETTA, G. E. CANTARELLA, AND M. DI GANGI

A within-day and day-to-day dynamic assignment model for a general network has been proposed recently. The model follows a nonequilibrium approach, in which flow fluctuations are modeled as a stochastic process. It includes a model of dynamic network loading for computing within-day variable arc flows from path flows. In this paper, the sensitivity and the operational characteristics of the model are tested by analyzing some effects of control measures on a small realistic network. The results of these applications show that the proposed model is a valid tool to estimate the effectiveness of some traffic engineering measures and informative systems. It also appears that some control measures cannot be assessed correctly without the explicit simulation of the demand elasticity over departure times and of the day-to-day adjustment process determined by users' memory and forecasting.

Recently, the stochastic process approach to the analysis of transportation system dynamics (*1,2*) has been extended to account for both within-day and day-to-day temporal fluctuations of demand and flows (*3*). Following this approach the evolution of the system over time is analyzed rather than seeking an equilibrium or self-reproducing solution (if any), as in within-day constant (*4,5*) and within-day dynamic equilibrium models. The latter can be developed on the basis of deterministic (*6,7*) or stochastic (*8,9*) users' behavior models.

The stochastic doubly dynamic model, described in this paper, allows the simulation of system adjustments following network modifications, the role of habit in users' choices, and the effects of some informative systems and control strategies. In addition, the model can be extended to cover the case of real-time informative strategies, allowing users to change their path en route.

In this paper, the general structure of the model is briefly outlined. Potential applications to a small real-size network are then presented to show the effects of traffic engineering measures, different demand structures and levels, and different types of informative systems.

DYNAMIC DEMAND/SUPPLY INTERACTION MODEL

Day-to-day dynamics refers to system variations occurring between successive reference periods, which can be part of the day, (e.g., the morning peak period) or the whole day. In the following, the reference period will be called "day" and denoted by index "t".

Within-day dynamics refers to variations taking place within the day and is analyzed by dividing the day into n_h subperiods. In the following, the generic subperiod will be called "interval" and denoted by index "h"; with no loss of generality, the interval length will be assumed constant and equal to T units of time.

Definitions and Notations

The transportation system is represented by a network; the generic origin-destination (O-D) pair is denoted by i, the generic path of the network by k, and the set of indexes relative to paths connecting the O-D pair i by K_i.

The total number of users deciding whether, when, and how to travel between each O-D pair i is denoted by d_i; it is assumed to be known and constant in each day t. Elasticity of the demand level; that is, changes in the number of users traveling each day, can be simulated by defining a fictitious interval ($h = n_h + 1$) corresponding to the choice of not moving at all. The disutility associated with this alternative is defined by the utility of not moving at all.

Let F^t_{hk} be the flow of users following path k and leaving during interval h of day t. These values, arranged in a vector $\mathbf{F}^t$, are assumed to define the state of the system at day t. Demand and path flows are related, since

$$d_i = \sum_h \sum_{k \in K_i} F^t_{hk} \tag{1}$$

If $\bar{p}^t(h,k)$ denotes the fraction of users leaving during interval h of day t and following path $k \in K_i$ between the O-D pair i, the path flow can be expressed as:

$$F^t_{hk} = d_i \cdot \bar{p}^t(h,k) \qquad k \in K_i \tag{2}$$

An obvious simplification occurs if the demand temporal profile or the departure time fractions $\bar{p}_i(h)$ are fixed. In this case, the choice fractions can be expressed as $\bar{p}^t(h,k) = \bar{p}_i(h) \cdot \bar{p}^t(k/h)$, $k \in K_i$. Elasticity of the demand level; that is, changes in the number of users traveling each day, can be simulated also in this case, by introducing a fictitious path ($k = 0$) between each O-D pair to simulate, though only descriptively, the choice of not moving at all.

Let f^t_{ah} be the flow on arc a during interval h of day t, and $\mathbf{f}^t$ the corresponding arc flow vector. In a within-day dynamic

Dipartimento di Ingegneria dei Trasporti, Università degli Studi di Napoli, FEDERICO II, Via Claudio, 21-80125, Napoli, Italy.

framework, the definition of arc flow is not unique because time and space averages do not coincide any longer. The operative definition of arc flow then depends on the modeling framework adopted for network loading, as reported in the section on the model of users' behavior.

The correspondence between arc flows, however defined, and path flows (network loading mapping) can be formally expressed as

$$f^t = f^t(\mathbf{F}^t) \tag{3}$$

Let C^t_{hk} be the average generalized travel cost on path k leaving during interval h of day t, and $\mathbf{C}^t$ be the corresponding travel cost vector. Path travel costs are generally functions of the arc travel costs, which in turn are functions of arc flow vector; therefore, the following formal relationship holds:

$$\mathbf{C}^t = \mathbf{C}^t(f^t) \tag{4}$$

Strictly speaking, travel costs are random variables with average values that can be expressed as a function of arc flows by means of arc cost functions. The probability of the occurrence of a given cost vector conditional to the path flow vector can thus be expressed by Equation 3 as

$$Pr[\mathbf{C}^t/f^t] = Pr[\mathbf{C}^t/f^t(\mathbf{F}^t)] \tag{5}$$

On the other hand, most assignment models ignore the dispersion of travel times around their mean values; in this case, the conditional probabilities in Equation 5 can be substituted by the usual expression

$$\mathbf{C}^t = \mathbf{C}^t(\mathbf{F}^t) \tag{6}$$

obtained by combining Equations 3 and 4.

Stochastic Process Model

It appears realistic to assume that the number of users F^t_{hk} is an integer. Then, the number of feasible states, that is, path flow vectors with nonnegative components and consistent with the demand as defined by Equation 1, is finite.

For the dispersion of users' behavior and the intrisic randomness of some parameters (number of users, network conditions, travel costs, times, etc.), it is assumed that the system takes different states in different days. Furthermore, these states cannot be exactly forecasted by the analyst.

In other words, the actual values of fractions $\bar{p}^t(h,k)$ for a given day cannot be predicted in advance. Therefore, the evolution of the system among feasible states in successive days can be described as a stochastic process, with properties depending on the hypotheses made on users' behavior and network configuration. Interval/path fractions are realizations of random variables, whose average values are the choice probabilities $p^t(h,k)$, which can be obtained by a properly defined model (see the next section for an example).

The probability $Pr[\mathbf{F}^t]$ that the system is in a given state $\mathbf{F}^t$ at day t can be computed, at least theoretically, from choice probabilities:

$$Pr[\mathbf{F}^t] = Pr[\mathbf{F}^t/p^t(h,k)\ \forall h,k] \tag{7}$$

It can be reasonably assumed that users choose paths and departure times using information about times and costs that have occurred in previous days $\mathbf{C}^{t-1}$, $\mathbf{C}^{t-2}$, . . . , either because this is the only available information or because it complements information supplied by an informative system.

If some travelers choose using a real-time informative system, choice probabilities depend also on times and costs incurred in the current day. More precisely, in the case of a real-time trip planning system, departure time and path probabilities depend on costs incurred at most up to the departure interval, in the case of a real-time route guidance system, path choice probabilities depend on costs up to the arrival time (computed from the departure time and the travel time). Similar conditions occur if an adaptive route choice behavior is assumed for users. Then the following will result:

$$p^t(h,k) = p^t(h,k)[\mathbf{C}^t, \mathbf{C}^{t-1}, \mathbf{C}^{t-2}, \ldots]$$

Moreover, if it is assumed that users have a limited memory, that is they are significantly influenced in their choices at most by a limited number (m) of past days, then the following results are obtained:

$$p^t(hk) = p^t(h,k)[\mathbf{C}^{t-i}, i = 0, \ldots, m] \tag{8}$$

Since path travel times and costs in congested networks depend deterministically or stochastically on arc flows, it turns out that the probability that the system is in a given state $\mathbf{F}^t$ at day t depends on the states occupied by the system in m previous days. Combining Equations 5, 7, and 8

$$\begin{aligned} &Pr[\mathbf{F}^t/\mathbf{F}^t, \mathbf{F}^{t-1}, \ldots, \mathbf{F}^{t-m}] \\ &= Pr[\mathbf{F}^t/p^t(h,k)[\mathbf{C}^{t-i}, i = 0, \ldots, m]] \cdot Pr[\mathbf{C}^t/\mathbf{F}^t] \\ &\cdot Pr[\mathbf{C}^{t-1}/\mathbf{F}^{t-1}] \cdot \ldots \cdot Pr[\mathbf{C}^{t-m}/\mathbf{F}^{t-m}] \end{aligned} \tag{9}$$

If path costs are assumed not to be random variables, then

$$\begin{aligned} &Pr[\mathbf{F}^t/\mathbf{F}^{t-1}, \ldots, \mathbf{F}^{t-m}] \\ &= Pr\{\mathbf{F}^t/p^t(hk)[\mathbf{C}^{t-1}(\mathbf{F}^{t-1}), \quad 0 = 1, \ldots, m]\} \end{aligned}$$

It can be proven, by using results of m-dependent Markov chains (*2*) that the process admits a unique stationary probability distribution and it is ergotic if, in addition to the limited memory assumption (Equation 8), the following (sufficient) conditions hold:

1. Choice probabilities, given the same sequence of costs relative to the previous day, and possibly to the same day, are time homogeneous, that is, invariant with respect to a time translation:

$$p^t(hk)[\mathbf{C}^{t-i}, i = 0, \ldots, m] = p^{t'}(hk)[\mathbf{C}^{t'-i}, i = 0, \ldots, m]$$
$$\text{if } \mathbf{C}^{t-i} = \mathbf{C}^{t'-i}, i = 0, \ldots, m$$

2. For each pair of different states there exists at least one sequence of feasible states, with strictly positive transaction probabilities, from one to the other.

The stated conditions depend only on the assumption that behavioral rules are constant over time, and do not depend on any assumptions about the particular type of users' choice behavior and the information available to them, apart from that of finite memory. In other words, different types of choice models can be used for departure time and path choice, such as random utility or noncompensatory models in so far they do not assume that users' experience influence their behavior thereafter.

It is also worth noting that random events modifying network performances, such as accidents and bad weather conditions, can be included in the proposed framework if their occurrence probability law is stable over time.

Because of the existence and unicity of the stationary distribution, one distribution of path flows can be associated to each demand-supply system, independent of the starting configuration. Process ergodicity allows the computation of flow means and moments through the simulation of only one realization of the process. It is worth noting that the mentioned properties are satisfied regardless of the type of arc cost functions.

Obviously it is still possible to study transitions between two stationary states of the system. In this case, the ergodicity property no longer applies, and flow moments must be computed over repeated simulations of transients.

The stationary probability distribution of flows could have different modes, denoting a situation comparable to that of multiple equilibria. However, in the proposed stochastic process approach, the whole probability distribution for each arc flow can in principle be computed, although in the multiple equilibria case no method known to the authors guarantees information on all equilibrium configurations.

Solution Approach

In the following, a solution approach for computing expected values and moments of time-varying arc flows, both in steady state and transient conditions, will be described briefly.

Full specification of the assignment model requires a modeling of departure time and path choice, of users' learning and forecasting mechanisms, and possibly of the informative strategy and users' reactions. Most models proposed in the literature to simulate the aspects given previously could be adapted to fit into the proposed framework. For instance, departure time choice could be simulated by a random utility model (*10*) or by the "bounded rationality" model proposed by Mahmassani and Chang (*11,12*). A possible specification will be described in the next section.

The number of users choosing each departure interval/path alternative in a given day can be obtained from choice probabilities by using a Monte Carlo simulation or by substituting probabilities to fractions. In the latter case, the resulting sequence is a pseudo realization of a stochastic process. An obvious simplification occurs when fractions $p_i^t(h)$ are exogenously given (prefixed demand profile).

Once that choice fractions and, consequently, path flows are known for the current day, arc flows can be computed by a dynamic network loading method, as described in a following section. Arc flows allow the computation of travel times and thus the update of travelers' information and forecasts to be used for modeling choice probabilities of the next day.

Flows for each arc of the network and for each interval in the reference period can be used for two purposes. The first is to estimate stationary means and moments for arc flows; the second is to estimate means and moments during transients caused by any modification in the network, in the demand, or in both the network and the demand.

The reaching of stationarity can be checked by performing a Student's-*t* test on the differences between average arc flows in each interval over two successive sequences of days.

MODEL OF USERS' BEHAVIOR

Users' Choice Behavior

Users' choice behavior has been modeled by the random utility model proposed by Small (*10*) and reformulated by Ben Akiva et al. (*13*). This model has been slightly modified to explicitly introduce a "habit effect" in users' choices. In particular it is assumed that each day only a fraction Ω of users reconsider the previous days' choice, and that they give an extra utility to the alternative chosen the previous day.

It is assumed that each user deciding how (path k) and when (departure interval h) to travel at day t associates to each alternative (h,k) a perceived utility expressed by the sum of a systematic utility and a random residual error.

The systematic utility represents the average predicted utility, whereas the random residual takes into account different perception errors made by users (e.g., relative to travel times and costs) and the dispersion of some characteristics within the population of decision makers (desired arrival times, reciprocal substitution coefficients, missing attributes, etc.).

The systematic utility $\overline{V}_{hk}^t$ can be expressed using a modified version of the model proposed by Small (*10*) as the sum of disutilities relative to the generalized transportation cost and to early or late arrival penalty:

$$\begin{aligned}\overline{V}_{hk}^t = &-(1 - \mu_{hk}) \cdot (\beta_{1i}\overline{C}_{hk}^t \\ &+ \mathrm{MAX}\{\beta_{2i}[(D_i - \delta_{1i}) - \overline{B}_{hk}^t], 0\} \\ &+ \mathrm{MAX}\{\beta_{3i}[\overline{B}_{hk}^t - (D_i + \delta_{2i})], 0\})\end{aligned} \tag{10}$$

μ_{hk} = $\mu\varepsilon[0,1]$ if (h,k) is the alternative chosen the previous day,
= 0, otherwise this parameter reduces proportionally the disutility for the choice of the same path and departure time as in the previous day $t - 1$, and attempts to capture the conservative behavior of users;

$\overline{C}_{hk}^t$ = average predicted generalized transportation cost along path k starting during interval h, on day t;

$\overline{B}_{hk}^t$ = average predicted arrival time, starting during interval h and moving along path k, on day t, computed as $(h - 1) \cdot T + \overline{T}_{hk}^t$;

T_{hk}^t = average predicted travel time along path k starting during interval h, on day t;

D_i = desired arrival time, variable with the O-D pair (and category) i; and

$\beta_{1i}, \beta_{2i}, \beta_{3i}$ = reciprocal substitution coefficients, variable with the O-D pair (and category) i.

This expression assumes that users of O-D pair i have a tolerance interval $[-\delta_{1i}, \delta_{2i}]$ around the desired arrival time, and early or late arrivals cause a disutility proportional to the advance or the delay.

On day t average values of predicted travel time and generalized transportation cost of path k leaving during interval h can be computed through duly defined filters, which model the learning and forecasting mechanisms used by the average traveler, including the interaction with an informative system, if any. Different filters can be used for different kinds of users, for example, commuters versus noncommuters, informed versus noninformed, and so on.

Systematic utility has been defined assuming the generalized transportation cost to be equal to the travel time. The average perceived travel time has been computed as a weighted average of the previous day actual travel time T_{hk}^{t-1} and of the previous day average perceived travel time $\overline{T_{hk}^{t-1}}$:

$$\overline{T}_{hk}^{t} = \tau \cdot \sum_{i=1}^{t-1} (1-\tau)^{i-1} \cdot T_{hk}^{t-i} + (1-\tau)^{t} \cdot T_{hk}^{0}$$

$$= \tau \cdot T_{hk}^{t-1} + (1-\tau) \cdot \overline{T_{hk}^{t-1}} \qquad (11)$$

where $\overline{T}_{hk}^{0} = T_{hk}^{0}$ is a starting value. Values of τ close to one denote a stronger influence of the previous day travel time. Similar filters have been proposed by Mahmassani and Chang (*12*) and Iida et al. (*14*).

It has been assumed that each day a prefixed fraction of users Ω takes into consideration the possibility of modifying their previous day choice (but they do not necessarily have to). The choice probabilities of the users that reconsider their choice is simulated through a path/departure time nested logit model:

$$p^t(h,k) = p^t(h) \cdot p^t(k/h) \qquad (12)$$

$$p^t(k/h) = \exp[\Theta_1 \overline{V}_{hk}^t] / \sum_j \exp[\Theta_1 \overline{V}_{hj}^t] \qquad (13)$$

$$p^t(h) = \exp[\Theta_2 \overline{Y}_h^t] / \sum_j \exp[\Theta_2 \overline{Y}_j^t] \qquad (14)$$

where

Θ_1 = Weibull-Gumble parameter of the random residual relative to the pair (h,k),
Θ_2 = $(1/\Theta_1^2 + 1/\Theta^2)^{(-1/2)}$,
Θ = Weibull-Gumble parameter of the random residual relative to the interval h only, and
Y_h^t = $(1/\Theta_1) \cdot \ln \Sigma_j \exp[\Theta_1 V_{hj}^t]$ is the logsum variable relative to interval h.

If $\Theta_1 = \Theta_2$ the above model reduces to the factorialization of a multinomial logit over the pair (h,k).

As it is known from the logit theory, coefficients Θ_i are inversely proportional to the standard deviation σ_i of the perception error in path and departure time choice, respectively. For each O-D pair values of coefficients Θ_i have been computed by assuming a prefixed value of the variation coefficient Cv_i for each of them:

$$\Theta_1 = \Pi/(\sqrt{6} \cdot \sigma_i) \cong 1.282/\sigma_i = 1.282/(Cv_i \cdot \overline{V}) \qquad (15)$$

where $\overline{V}$ is the value of utility obtained by averaging across all the alternatives, the average perceived utility as given by Equation 10.

Thus different quality in information can be simulated by differentiating all the users of each O-D pair in two or more types (e.g., informed and not informed) with different variation coefficients, and consequently, different values of Θ_i.

Users' Behavior and Informative Systems

Generally users moving between an O-D pair are assumed to choose at their origin the departure interval and a path k_0 to reach their destination. After leaving the origin, rerouting during the trip may occur because of adaptive behavior at duly located diversion nodes (possibly at each node of the network). Therefore, the actually used path k may be different from the initially chosen k_0. The choice probability at day t, $p^t(h,k)$, can be expressed as

$$p^t(h,k) = p_i^t(h) \cdot p^t(k/h) \qquad k \in K_i$$

where

$p_i^t(h)$ = probability of choosing departure interval h for a user traveling between O-D pair i; and
$p^t(k/h)$ = probability to follow path $k \in K_i$ from the origin to the destination of O-D pair i, once departure interval h has been chosen (it can be expressed as the joint probability that the sequence of paths from the origin to the first diversion node, between each pair of successive diversion nodes, and from the last diversion node to the destination forms path k).

An informative system may supply information or indications to users before they start (pretrip) or while they are traveling (en route). Moreover, the informative system can use exclusively information about the network conditions in the previous days (static) or combine them with information about the network conditions that occurred during day t (dynamic or real time) before the departure time (pretrip) or the arrival time (en route). Different classes of users with different types of available information can be taken into account.

The behavior of users not advised by an informative system can be modeled through usually adopted users' choice behavior models, as the one described in the previous subsection. The same kind of models, with a different parameter specification, can be adopted to deal with users advised by a static pretrip informative system (an example is given in a subsequent section).

A slight modification is needed in the case of a real-time pretrip informative system: Let

$$p_i^t[j = h] = p_i^t(h)$$

be the choice probability that begins the trip between O-D pair i during interval h of day t;

$$p_i^t[j \geq h] \begin{cases} = 1 & \text{if } h = 1 \\ = 1 - \sum_{m=1}^{h-1} p_i^t[j = m] & \text{if } h > 1 \end{cases}$$

be the choice probability that a user begins the trip during interval h or later; and

$$p_i^t[j = h/j \geq h] = p_i^t(h/\{h, h + 1, \ldots\})$$

be the choice probability that a user begins the trip during interval h conditional to leaving not earlier than interval h. This can be computed using a choice behavior model (as the one described in the previous section) assuming as choice set $\{h, h + 1, \ldots\}$.

Since

$$p_i^t[j = h/j \geq h] \cdot p_i^t[j \geq h] = p_i^t[j = h \cap j \geq h] = p_i^t[j = h]$$

the choice probabilities $p^t(h)$ can be recursively computed as

$$p_i^t(h) \begin{cases} = p_i^t(1/\{1, 2, \ldots\}) & \text{if } h = 1 \\ = p_i^t(h/\{h, h+1, \ldots\}) \cdot \left(1 - \sum_{m=1}^{h-1} p_i^t(m)\right) & \text{if } h > 1 \end{cases}$$

Once choice probability $p_i^t(h)$ has been computed, the path choice probabilities $p^t(k/h)$ can be computed through a users' behavior model. The resulting stochastic process maintains all the previously mentioned stationarity and ergodicity properties.

If a static or dynamic en route informative (or route guidance) system is operating for some users, the choice probabilities for the departure interval h and the initially chosen path k_0 can be computed as described previously. Then during the dynamic network loading stage at duly located diversion nodes (beacons of the route guidance system, or each node of the network), users are allowed to reroute onto a new path, and so on, until the destination is reached, according to a behavior model and the information or indications supplied by the route guidance system (as better explained in the next section).

A DYNAMIC NETWORK LOADING METHOD

In this section, a model for dynamic network loading; that is, computation of time-varying arc flows from a given path flow pattern in a given day t, is described and compared with other state-of-art models. A solution algorithm is also presented (*3*). In the following, superscript t will be omitted for the sake of simplicity.

Notation

Notations used in this section are as follows. Vectors and arrays are not explicitly cited.

k = path,
a = arc,
a' = arc (if any) following arc a on path k,
h = interval,
j = interval,
(j,k) = group leaving during interval j and traveling on path k,
f_{ah} = arc flow,
F_{jk} = path flow,
l_a = running arc length,
q_{ah} = queue at the end of interval,
Q_{ah} = capacity,
s_{ah}^{jk} = abscissa on running arc,
$u_{ah} = \overline{u}_{ah}$ = undersaturation delay,
$v_{ah} = V_{ah}(f_h)$ = running speed,
w_{ah}^{jk} = waiting time for queuing arc,
$y_{jk \atop ah}$ = arrival time at arc a,
$y_{jk \atop a'h}$ = arrival time at arc a',
$z_{jk \atop ah} = z_{ah}(y_{jk \atop ah})$ = oversaturation delay, and
$\alpha_{jk \atop ah}$ = crossing fraction.

Statement of the Problem

Previously it was stated that the relationship between arc and path flows in a within-day dynamic context is not trivial. It was also observed that arc flows are not even uniquely defined in this case. It is still possible to express the relationship between path and arc flows, however defined, in a way that is formally similar to the within-day uniform case.

Denoted by $\alpha_{ah}^{jk} \varepsilon [0,1]$, the fraction of *path flow* F_{jk} contributing to *arc flow* f_{ah}, named *crossing fraction*, the flow on arc a can be expressed by

$$f_{ah} = \sum_{jk} \alpha_{ah}^{jk} F_{jk} \tag{16}$$

If arc a does not belong to path k, or the starting interval of flow F_{jk} follows interval $h(j > h)$, or flow F_{jk} does not occupy arc a during interval h, the fraction α_{ah}^{jk} is equal to zero.

Crossing fractions can be arranged in matrices such as $A_{hj} = \{\alpha_{ah}^{jk}\}_a^k$, arc-path crossing matrix between arc-flows during interval h and path flows leaving in interval j. Denoted by f_h, the arc flow vector during interval h and by F_j, the path flow vector leaving during interval j, Equation 16 can be stated in matrix form as follows:

$$f_h = \sum_{j=1}^{h} A_{hj} \cdot F_j \tag{17}$$

For the whole day Equation 16 can be expressed as

$$f = A \cdot F \tag{18}$$

where

f = arc flow vector for the whole day,
F = path flow vector for the whole day, and
A = arc-path crossing matrix for the whole day formed by matrices A_{hj} with $A_{hj} = O$ (if $h < j$, A is a low triangular block matrix)

It is worth noting that the arc-path crossing matrix A is a generalization of the usual arc-path incidence matrix, because the following should result:

$$\sum_{m=1}^{nh} \alpha_{am}^{jk} = 1 \tag{19}$$

if all the users entering the network during interval j leave it in some interval $h \geq j$.

Generally, crossing fractions α_{ah}^{jk} depend on the definition adopted for arc flows, the network topology, and the time needed to reach arc a traveling on path k. Therefore, they depend on the travel times on arcs preceding arc a along path k, which in congested networks are function of the arc flows. Hence, it generally results that

$$A = A(f) \tag{20}$$

Therefore combining Equations 18 and 20 the following fixed-point problem is obtained:

$$f^* = A(f^*) \cdot F \tag{21}$$

A dynamic network loading method is essentially an algorithmic definition of crossing fractions, that is of relationship $\alpha_{ah}^{jk} = \alpha_{ah}^{jk}(f)$ needed to solve Equation 21.

It is worth noting that in noncongested networks in which travel times and delays are constant and independent of arc flows, the arc-path crossing matrix does not depend on arc flows and Equation 21 reduces to $f^* = AF$, as in the case of within-day constant demand.

Several methods have been proposed to solve the dynamic network loading problem, which in any case could be solved by a discrete simulation technique (microsimulation), although at the expense of a large computational effort. Generally, these methods give different results depending on the different hypotheses adopted.

Some methods do not explicitly formulate and solve the fixed-point problem (Equation 21). They are based on a generalization of network loading procedures used in static deterministic user equilibrium (*6,7*). A different method to indirectly solve the problem (Equation 21) has been proposed recently by Vythoulkas (*9*). This model is based on the discretization of a differential equation for each arc, expressing the relationship between the time derivative of the number of users on the arc and the difference between inflow and outflow.

All the preceding methods rely on assumptions that do not rule out some counterintuitive results such as overtaking between vehicles traveling on the same path and vehicles leaving in different times. In addition, these methods can hardly be extended to include en route diversions from the initially chosen path, because of the computational burden of keeping the identity of diverted path flows.

In the following, a new method for dynamic network loading is described, which directly solves the fixed-point problem (Equation 21), thus overcoming some of the drawbacks of the other proposed methods.

General Hypotheses and Definitions

The set of all users leaving in the same interval j and following the same path k is called *group* or packet (j,k). All users of a group are assumed to experience the same trip as the group leader, whose departure occurs at a prefixed instant (the middle or the beginning) of the interval. Hence, if an arc is occupied by the leader of a group during an interval, it is occupied by all the users belonging to that group (*grouping hypothesis*).

This assumption appears acceptable for usual O-D demand flows and interval lengths. In any case its realism can be improved by reducing the interval duration, or by subdividing a group into smaller units. Currently, an enhanced model for dynamic network loading relaxing this hypothesis is being developed by the authors.

In the following it is also assumed, for simplicity, that a user group follows the path chosen before starting the trip until its destination is reached. However, the described method can be easily extended to include while-trip rerouting. In this case, at duly located diversion nodes (eventually each node can be a diversion node), group (j,k) traveling on path k may reroute on a new path k' from the diversion node to the destination, thus becoming group $(j,k - k')$. This can be the result of a simple adaptive behavior or of the interaction with a route guidance system, providing indications or information. In both cases, a choice model is needed to represented users' reactions resulting in a change from path k to path k'.

Two types of arcs requiring different modeling approaches will be considered in the following:

- *Running arcs* (e.g., a stretch of a street): for which the time needed to leave the arc is continuously spent along its length; and
- *Queuing or waiting arcs* or bottlenecks (e.g., a junction approach): for which delay occurs only at the end of the arc, assumed of null length, because of queuing due to capacity constraints.

Obviously the simulated network can include both types of arc.

Let v_{ah} be the *average running speed* on running arc a, with length l_a, during interval h. Running speed is assumed to be the same for all users traveling on the arc during interval h (*equal running speed hypothesis*), regardless of when they have entered the arc.

Let u_{ah} be the *undersaturation delay* for queuing arc a, during interval h. Undersaturation delay is assumed to be the same for all users entering the arc during interval h (*equal undersaturation delay hypothesis*) regardless of when they have entered the arc.

Let z_{ah}^{jk} be the *oversaturation delay* for group (j,k) joining the queue at bottleneck arc a during interval h. It is assumed to be equal for all the users of group (j,k) (*group specific oversaturation delay hypothesis*) and depending on the arrival time of group (j,k) at arc a during interval h. For undersaturated conditions, the oversaturation delay is equal to zero.

Group Movements on the Network

Let $y_{ah}^{jk} \in [0,T]$ be the *arrival time* of group (j,k) at arc a during interval h. It is meaningfully defined only if arc a belongs to path k and interval j precedes interval h, that is $j \leq h$ (otherwise it is assumed equal to zero by convention). In the following it is always assumed that arc a belongs to path k, and interval j precedes interval h, that is $j \leq h$. Let a' be the arc (if any) following arc a on path k. Naturally the exit time from arc a of group (j,k) is equal to the arrival time at arc a'.

If group (j,k) has not yet reached arc a during interval h, $y_{ah}^{jk} = 0$; vice versa if group (j,k) has already left arc a before interval h, $y_{ah}^{jk} = T$. Moreover, it is assumed that the arrival time is equal to zero for the first arc occupied during interval h.

The difference, $T - y_{ah}^{jk}$, between the interval length T and the arrival time y_{ah}^{jk} at arc a is the time still available to group (j,k) to move on arc a during interval h.

A group (j,k) arriving at a *running arc* a during interval h needs a time l_a/v_{ah} to entirely cover the arc. It can be proven that the exit time from arc a [or the arrival time at the next arc a' (if any) on path k)] and the abscissa reached on arc a by group (j,k) by the end of interval h are given by (also shown in Figure 1)

$$y_{a'h}^{jk} = \mathrm{MIN}[T, y_{ah}^{jk} + (l_a - s_{ah-1}^{jk})/v_{ah}] \quad (22)$$

$$s_{ah}^{jk} = \mathrm{MIN}[l_a, s_{ah-1}^{jk} + (T - y_{ah}^{jk}) \cdot v_{ah}] \quad (23)$$

assuming that

- $y_{ah}^{jk} = 0$ if arc a is the first occupied by group (j,k) during interval h;
- $y_{ah}^{jk} = 0$ if group (j,k) has left from arc a before interval h;
- $y_{ah}^{jk} = T$ if group (j,k) has not yet reached arc a by the end of interval h;
- $s_{a0}^{jk} = 0$
- $s_{ah}^{jk} = 0$ if group (j,k) has not yet reached arc a by the end of interval h; and
- $s_{ah}^{jk} = l_a$ if group (j,k) has left from arc a by the end of interval h.

If group (j,k) enters *queuing arc* a during interval h, the time needed to exit the arc is equal to the total delay: $u_{ah} + z_{ah}^{jk}$, whereas the time still available to move is given by the difference $T - y_{ah}^{jk}$.

If $u_{ah} + z_{ah}^{jk} \leq T - y_{ah}^{jk}$, group (j,k) leaves arc a during interval h and enters the next arc a' (if any) on path k at time

$$y_{a'h}^{jk} = y_{ah}^{jk} + u_{ah} + z_{ah}^{jk} \quad (24)$$

Vice versa, if $T - y_{ah}^{jk} < u_{ah} + z_{ah}^{jk}$, it remains on the arc. Let

$$w_{ah}^{jk} = u_{ah} + z_{ah}^{jk} - (T - y_{ah}^{jk}) > 0 \quad (25)$$

be the time needed by group (j,k) to leave arc a at the end of interval h (*residual waiting time*).

If group (j,k) is still on queuing arc a at the beginning of interval $h + 1$, it should stay queuing for a time w_{ah}^{jk}. If it results that $T \leq w_{ah}^{jk}$, the exit time from arc a for group (j,k) or the arrival time on the next arc a' (if any) on path k is given by

$$y_{a'h+1}^{jk} = w_{ah}^{jk} \quad (26)$$

otherwise it should stay on arc a for a time

$$w_{ah+1}^{jk} = w_{ah}^{jk} - T \quad (27)$$

If group (j,k) reaches its destination during interval h, it leaves the network and it turns out that $s_{ah}^{jk} = l_a$ or $w_{ah}^{jk} = 0$ for all arcs on path k, and $m \geq h$. Groups that do not reach their final destination by the end of the reference period can be left on the network without any loss of generality.

Computation of Crossing Fractions

To compute crossing fractions from running speeds and delays, an operational definition of arc flows has to be formulated.

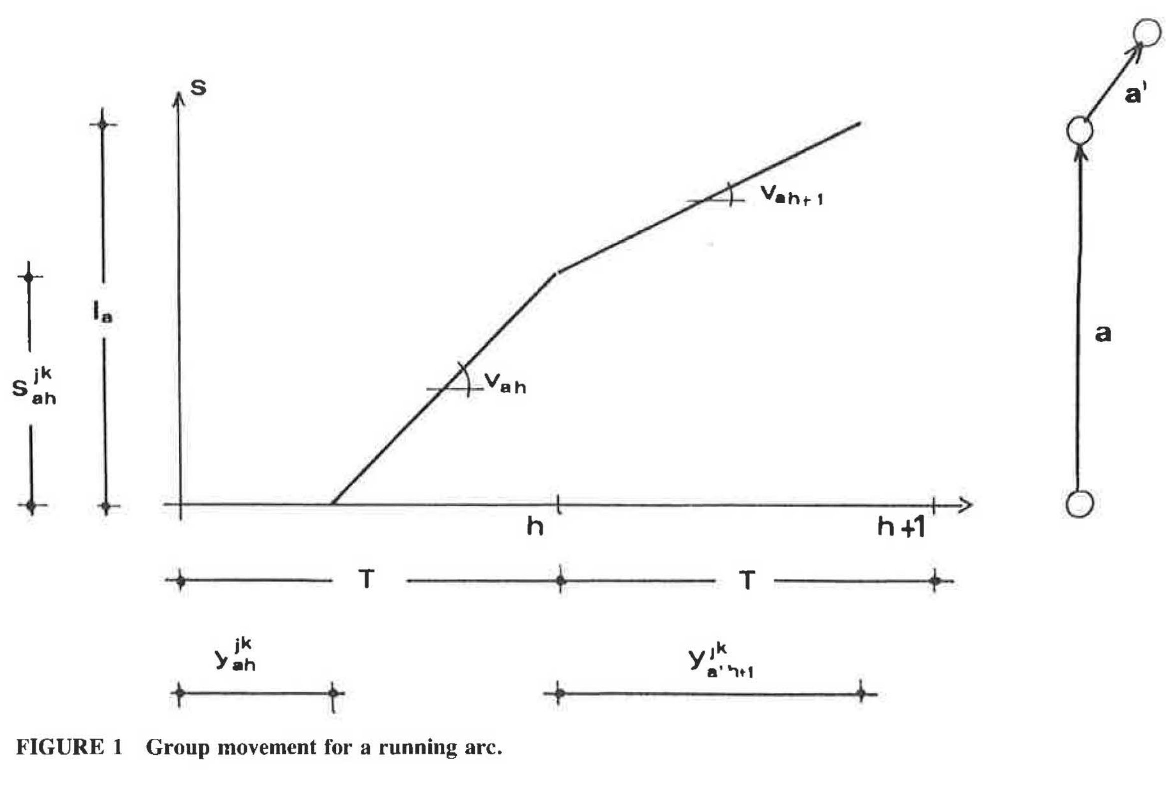


FIGURE 1 Group movement for a running arc.

For the grouping hypothesis, the average number of users per unit of time on *running arc a* during interval h is given by

$$\left(\sum_{jk} F_{jk} \cdot T \cdot \int_0^T \delta_{ah}^{jk}(t)\,dt\right) \Big/ T$$

Where

$\delta_{ah}^{jk}(t) = 1$ if group (j,k) is on arc a at time t; and
$\delta_{ah}^{jk}(t) = 0$ otherwise.

Moreover, for the equal running speed hypothesis, the time spent by group (j,k) on arc a during interval h is equal to the traveled length divided by the average running speed, that is

$$\int_0^T \delta_{ah}^{jk}(t)\,dt = (s_{ah}^{jk} - s_{ah-1}^{jk})/v_{ah}$$

Therefore, the average density on arc a during interval h, which is equal to the average number of users per unit of time divided by the arc length, is given by the expression $\sum_{jk} F_{jk} \cdot (s_{ah}^{jk} - s_{ah-1}^{jk})/(v_{ah} \cdot l_a)$

The flow f_{ah} on *running arc a* during interval h can be defined as the product of the average density and the average speed:

$$f_{ah} = \sum_{jk} F_{jk} \cdot (s_{ah}^{jk} - s_{ah-1}^{jk})/l_a$$

Therefore from Equation 16, it turns out that

$$\alpha_{ah}^{jk} = (s_{ah}^{jk} - s_{a(h-1)}^{jk})/l_a$$

Also from Equation 23 it turns out that

$$\alpha_{ah}^{jk}(s_{ah-1}^{jk}, v_{ah}, y_{ah}^{jk}) = \mathrm{MIN}\{(T - y_{ah}^{jk}) \cdot v_{ah}/l_a, (l - s_{ah-1}^{jk}/l_a)\} \qquad (28)$$

Moreover it results that

$$s_{ah}^{jk} = s_{ah-1}^{jk} + \alpha_{ah}^{jk} \cdot l_a \qquad (29)$$

Similarly for a *queuing arc*, the flow can be defined as the time average in flow. For the grouping hypothesis, it turns out that if group (j,k) reaches arc a during interval h, the corresponding crossing fraction is equal to one, it is equal to zero otherwise:

$$\alpha_{ah}^{jk}(y_{ah}^{jk}) \begin{cases} = 1 \text{ if } T - y_{ah}^{jk} > 0 \\ = 0 \text{ if } T - y_{ah}^{jk} = 0 \end{cases} \qquad (30)$$

For both types of arcs, if group (j,k) stops on arc a during interval h, it turns out that $\alpha_{ih}^{jk} = 0$ for any arc i following arc a on path k. If group (j,k) reaches its destination, it exits from the network and it turns out that $\alpha_{am}^{jk} = 0$, $m > h$, $a \in k$.

It can be proven that this relationship is satisfied by the preceding proposed definition of crossing fractions, for groups leaving the network by the end of the simulation period.

Solution Approach

In this section, the crossing fraction definition described in the previous section is used to build up a fixed-point formulation of the dynamic network loading problem, according to the considerations in section, "Statement of the Problem."

Summarizing the results of the preceding section, the crossing fraction of group (j,k) on a running arc depends on the running speed and the arrival time at that arc. On the other hand, for a queuing arc, the crossing fraction of group (j,k) depends on the arrival time at that arc.

In turn, according to next preceding section, the arrival time of group (j,k) at arc a can be sequentially computed for each arc of a given path from running speeds and delays, by Equations 22 through 27.

Hence crossing fraction α_{ah}^{jk} is a function of running speeds and delays on arcs preceding arc a on path k. In the case of a running arc it is also a function of the running speed on that arc, whereas if it is relative to a queuing arc it does not depend on the delay on the same arc. Then, it generally results that

$$\alpha_{ah}^{jk} = \alpha_{ah}^{jk}(s_{ah-1}^{jk}, w_{ah-1}^{jk}, \mathbf{v}_h, \mathbf{u}_h, \mathbf{z}_h) \qquad (31)$$

where $\mathbf{v}_h$, $\mathbf{u}_h$, and $\mathbf{z}_h$ are, respectively, the vectors of running speeds, undersaturation, and oversaturation delays occurred on the network during interval h.

Average running speed is assumed to be function of the vector of the arc flows during interval h, $\mathbf{f}_h$, through usually adopted cost-flow functions:

$$v_{ah} = v_{ah}(\mathbf{f}_h) \qquad (32)$$

The undersaturation delay is assumed flow-independent:

$$u_{ah} = \bar{u}_{ah} \qquad (33)$$

According to a fluid approximation deterministic model, it turns out that

$$z_{ah}^{jk} = z_{ah}(q_{ah-1}, f_{ah}, y_{ah}^{jk}) = \mathrm{MAX}[(q_{ah-1}/Q_{ah} + (f_{ah}/Q_{ah} - 1) \cdot y_{ah}^{jk}), 0] \qquad (34)$$

where

q_{ah} = *queue* on arc a at the end of interval h (assuming $q_{a0} = 0$);
Q_{ah} = *capacity* of arc a during interval h (it can vary in different intervals, for example in the case of a variable traffic light system).

Therefore, the expression for crossing fractions can be formally rewritten as

$$\alpha_{ah}^{jk} = \alpha_{ah}^{jk}(s_{ah-1}^{jk}, w_{ah-1}^{jk}, q_{ah-1}, \mathbf{f}_h) \qquad (35)$$

If crossing fractions are sequentially computed through this equation for each interval h, values s_{ah-1}^{jk}, w_{ah-1}^{jk}, and q_{ah-1} relative to previous intervals are known. In this case, therefore, the only unknown arguments are arc flows relative to interval h, and it results that

$$\alpha_{ah}^{jk} = \alpha_{ah}^{jk}(\mathbf{f}_h) \qquad (36)$$

Combining Equations 36 and 16, the fixed-point problem (Equation 21) is obtained, which can be decomposed in a sequence of fixed-point problems since A is a low-triangular block matrix:

$$f_h^* = \sum_{j=1}^{h} A_{hj}(f_{h'}^*) \cdot F_j \tag{37}$$

Each of the problems (Equation 37) can be solved by usual fixed-point methods, as described in the following algorithm:

$i: = 0; f^{(0)}: = f_0$

REPEAT

$i: = i + 1;$

$e^{(i)}: = \sum_{j=1}^{h} A_{hj}(f^{(i-1)}) F_j$

$f^{(i)}: = (\alpha + \beta/i) \cdot e^{(i)} + (1 - (\alpha + \beta/i)) \cdot f^{(i-1)}$

UNTIL $f^{(i)} - f^{(i-1)} \leq \varepsilon$

$f_h: = f^{(i)}$

where

f_0 = assigned arc flow pattern,
ε = assigned tolerance vector, and
α, β = parameters in the range [0,1].

It is worth noting that the results of the algorithm are not affected by the sequence in which groups are examined and loaded on the network.

Comparisons of this algorithm with alternative specifications and an analysis of effective values for parameters α and β are reported elsewhere (*15*). In the following, $\alpha = 0.10$ and $\beta = 0.00$ will be adopted, and the initial flow pattern is assumed equal to the average flow pattern over the previous days for the same interval.

NUMERICAL EXAMPLES

In this section some results relative to an application of the proposed procedure to a realistic network are described briefly. The test network refers to the town of Battipaglia, Italy, with about 30,000 inhabitants. Supply data are relative to the real network, global O-D demand has been generated through a simple gravity model, and choice behavior has been modeled by adapting literature models, as described previously.

This example aims to test the proposed procedure on a real case. For this reason no specific conclusions about the case studied or comparisons with observational data are reported.

Test Network

The network used to test the model is shown in Figure 2. It has 62 nodes, 168 arcs, 269 O-D pairs, and is connected by 891 paths. A total travel demand of 4,970 users was considered. The parameters adopted for the choice behavior model described previously are as follows:

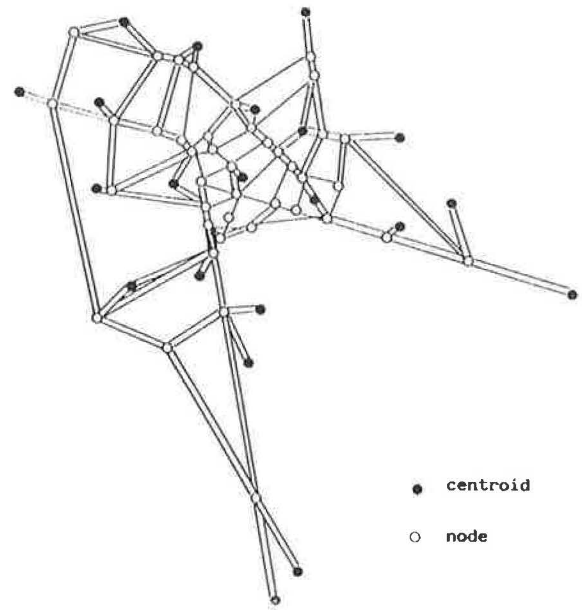


FIGURE 2 Test network.

Fraction of users reconsidering	*Parameter*
Their previous day choice	$\Omega = 0.5$
Reciprocal substitution coefficients	$\beta_1 = 1$, $\beta_2 = 1$, and $\beta_3 = 4$
Logit variation coefficients	$Cv_1 = Cv_2 = 0.20 \div 0.40$
Extra utility weight	$\mu = 0.10$
Filter parameter	$\tau = 0.90$

Davidson's cost functions were adopted for running arcs giving:

$$v_{ah}(f_{ah}) = v_a/[1 + 0.2 \cdot f_{ah}/(Q_{ah} - f_{ah})]$$

where v_a is the free-flow speed and Q_{ah} is the arc capacity. The value 0.2 is a calibration parameter that should be estimated by using actual data. For $f_{ah}/Q_{ah} \geq b_a$ (with b_a positive and less than one), the tangent approximation has been considered to avoid computational problems with asympotic functions. The greater the value b_a, the more sensitive to congestion the function is. In the following a value 0.80 is used, unless otherwise stated. For simplicity, no bottleneck was introduced in the network.

The simulation period lasts 60 min (the morning peak hour from 7:30 a.m. to 8:30 a.m.). It has been divided in 12 intervals with a length $T = 5$ min. Users are allowed to leave in the first 6 intervals (from 7:30 a.m. to 8:00 a.m.); the last 6 intervals have been included for system clearing—to allow all users to reach their final destination. A common decided arrival time has been assumed equal to 7:50, with $\delta_1 = 2.5$ and $\delta_2 = 2.5$.

The stationarity test was adopted on the basis of the comparison of the arc flow time means over two successive 10-day periods. Longer periods are not efficient, because they delay the time at which stationarity is recognized. On the other hand, using shorter periods the results of the test can be affected by periodic solutions.

Simulation of Traditional Control Strategies

The effects of some modifications in demand and supply have been simulated to evaluate the capabilities of the proposed model. Simulation scenarios were generated as follows:

- N1—an increase of total demand from 4,970 to 6,608; and
- N2—an increase of tolerance $\delta_1 = \delta_2 = 7.5$.

Case N1 is aimed at showing the effects of an increase of congestion, and case N2 represents a demand management measure (flexible work times).

The excess generalized cost (computed according to Equation 10) and travel time per user—difference between the actual and the zero-flow values—are compared with the results obtained without any modifications (STD) in Table 1, together with the average percentual changes in departure time and path demand patterns over successive days. Two different values of the variation coefficient of perception error (Equation 15), $Cv = 0.20 \div 0.40$, were used.

As expected, an increase of travel demand causes an increase of the generalized cost and travel time per user (case N1).

An increase of the tolerance band causes a significant reduction of generalized cost per user, and a smaller decrease of travel time. The first effect is a result of the reduction of late or early arrival penalty for users keeping their departure time, whereas the second one can be attributed to elasticity over departure times and a reduction of congestion (case N2), as shown in Figures 3 and 4, which report the departure (continuous line) and arrival (dashed line) profiles for cases N1 and N2.

Generally it seems that quite different values of the coefficient of variation of perception errors cause not great modifications of costs.

In all cases, the modifications in the demand structure with respect to the previous day values are very small, confirming the substantial stability of the day-to-day adjustment process adopted in spite of the quite high influence of the previous

TABLE 1 COMPARISON OF DEMAND MANAGEMENT STRATEGIES

	Excess General Cost (min)		Excess Travel time (min)		Changing users (fraction)	
Cv	0.20	0.40	0.20	0.40	0.20	0.40
STD	5.39	6.77	3.48	2.37	.002	.000
N1	6.94	8.36	3.40	3.65	.010	.000
N2	4.46	5.14	3.00	3.25	.000	.010

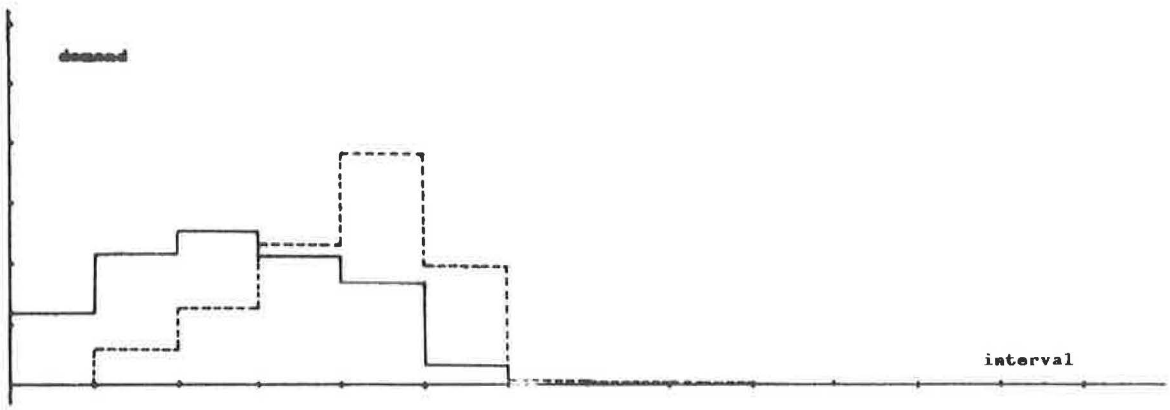


FIGURE 3 Departure and arrival profiles for case N1.

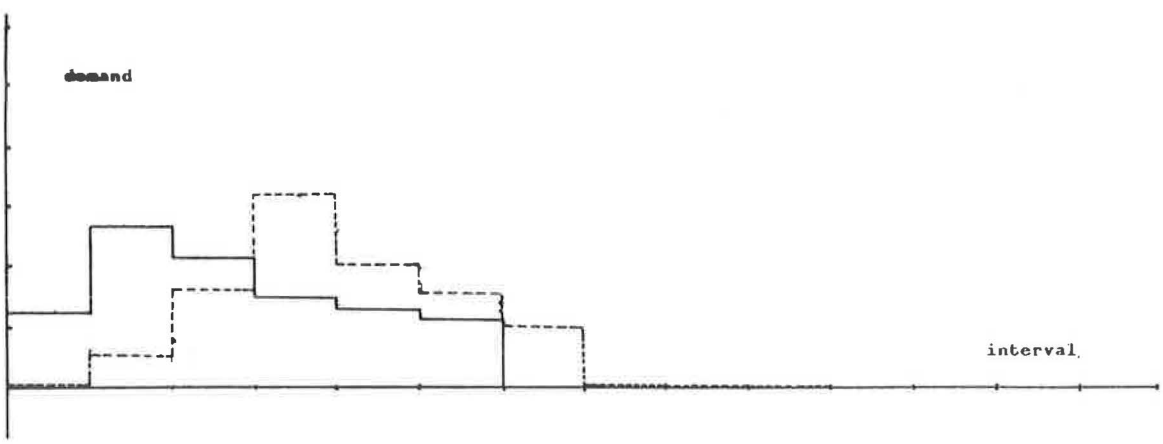


FIGURE 4 Departure and arrival profiles for case N2.

day information in the memory filter (τ = 0.9 in Equation 20).

Simulation of Informative Control Strategies

Using the travel demand value equal to 6,608, the effects of the introduction of a pretrip informative system based on historical data has been modeled. The drivers' reactions to the information provided was simulated by eliminating their inertia to change (the fraction Ω of users reconsidering the previous day choice was set equal to one, and no "habit" externality was considered, μ = 0) and by assuming a much lower dispersion in their choices with respect to the "predicted costs" given by the system, Cv = 0.05. Moreover it has been assumed that time and cost forecasts supplied by the informative system are less dependent on the recent past (filter parameter τ = 0.50 in Equation 20).

Some scenarios were examined considering different "market penetration" of the informative system:

- T1—1 percent of users are informed,
- T2—10 percent of users are informed,
- T3—50 percent of users are informed,
- T4—90 percent of users are informed, and
- T5—100 percent of users are informed.

The results are reported in Table 2, together with the results of case N1 as reference (*i* denotes informed users).

As expected, informed users experience lower generalized costs than noninformed ones (with a reduction of about 20 percent), and a small increase of travel time, since they have a better perception of the trade-off between travel time, even in congested intervals, and early or late arrival penalty. As an example, Figure 5 shows the departure and arrival profiles for case T2 for noninformed users (above) and informed users (below).

In addition, for more than 50 percent of informed users, generalized costs decrease for both types of users, as the fraction of informed users increases, leading to conditions which are better than the case N1. However, the difference between noninformed and informed users is not greatly affected by the fraction of informed users.

All these effects occur both for low (0.20) and high (0.40) values of the variation coefficient of the perception error of noninformed users; the value of this parameter does not affect the level of costs of informed users, a reduction of about 20 percent and 30 percent, respectively, occurs in comparison with noninformed users.

To show the sensitivity of the network to travel time function specification, the same scenarios have been simulated

TABLE 2 COMPARISON OF INFORMATIVE STRATEGIES

	Excess General cost (min)		Excess Travel time (min)		Changing users Fraction	
Cv	0.20	0.40	0.20	0.40	0.20	0.40
N1	6.94	8.36	3.40	3.65	.010	.000
T1	6.94	8.35	3.39	3.65	.010	.000
i[a]	5.45	5.51	3.72	3.97	.020	.000
T2	6.79	8.39	3.30	3.53	.005	.010
i	5.35	5.57	3.67	3.88	.018	.024
T3	6.37	7.67	2.93	2.93	.002	.000
i	4.88	4.91	3.27	3.33	.006	.000
T4	6.40	7.53	2.47	2.38	.000	.000
i	4.87	4.85	2.90	2.90	.006	.003
T5 *i*	6.51	6.51	2.65	2.65	.010	.010

NOTE: [a]*i* denotes informed users.

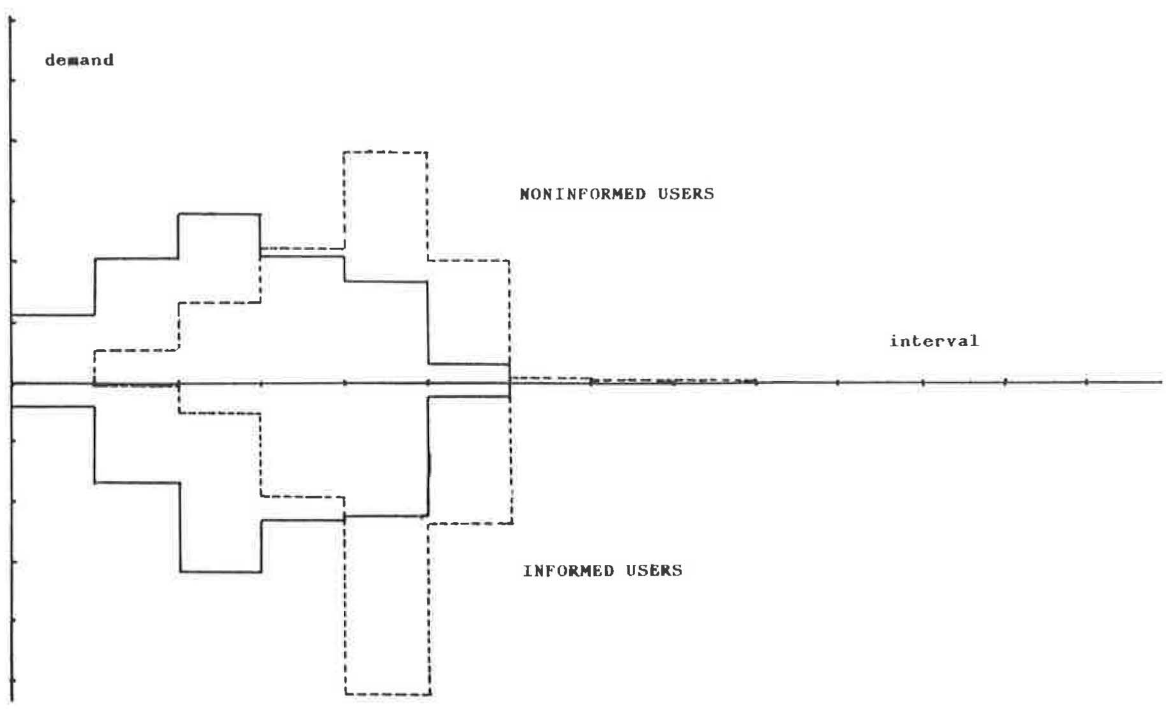


FIGURE 5 Departure and arrival profiles for case T2 (b_a = 0.80).

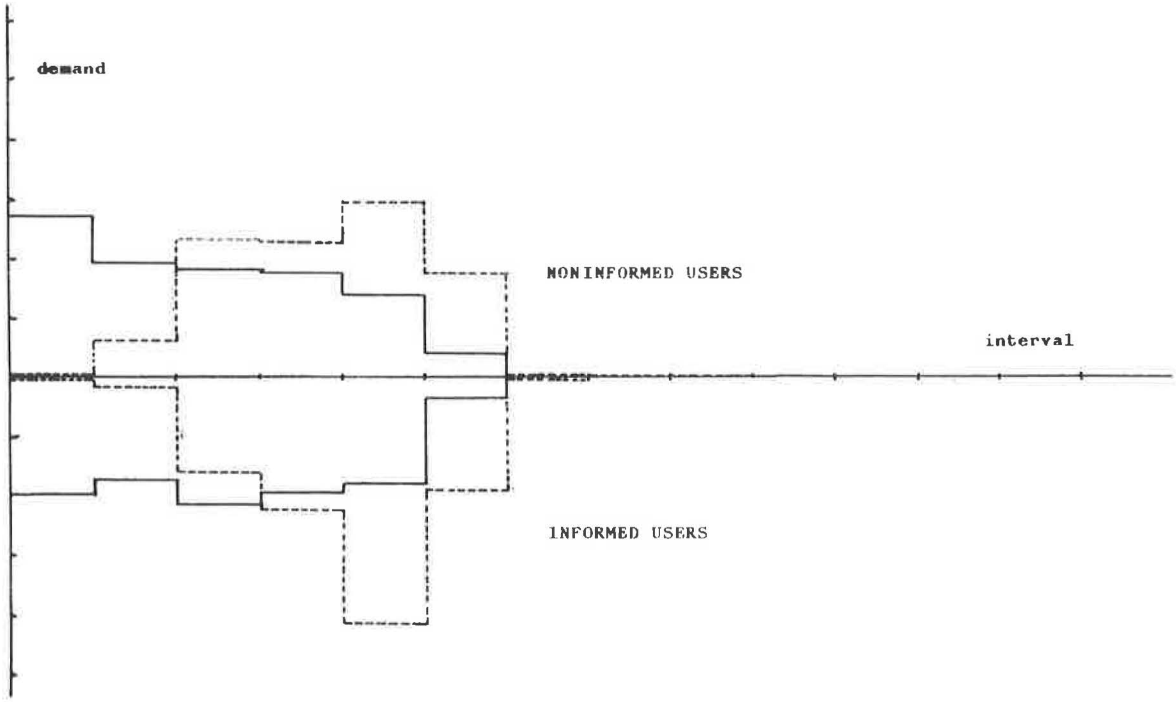


FIGURE 6 Departure and arrival profiles for case T2 (b_a = 0.90).

assuming a value $b_a = 0.90$ instead of 0.80, considering only the lowest value for the variation coefficient for noninformed users. These results are shown in Table 3. The results of the previously examined cases are generally confirmed. However, as a comparison with the results in Table 2, an increase occurs more in generalized cost than in travel time, since the values of travel times for arc flows close to capacity are higher and the users should spread their departure times to avoid congestion, thus less users can be on time. As an example, Figure 6 shows the departure and arrival profiles for case T2 for noninformed users (above) and informed users (below).

The preceding results, although inconclusive, indicate that informative control strategies may lead to better system conditions, in addition, as the fraction of informed users increases the performance of the system as a whole may become better both for informed and noninformed users.

TABLE 3 COMPARISON OF INFORMATIVE STRATEGIES

	Excess General cost (minutes)	Excess Travel time (minutes)	Changing users Fraction
Cv	0.20	0.20	0.20
N1	8.70	3.84	.010
T1	8.70	3.83	.010
i[a]	7.27	4.14	.015
T2	8.66	3.78	.000
i	7.24	4.10	.000
T3	9.07	3.48	.036
i	7.62	3.81	.074
T4	7.68	2.82	.021
i	6.04	3.41	.060
T5 *i*	6.08	3.38	.048

NOTE: [a]*i* denotes informed users

CONCLUSIONS

In this paper some applications of a model recently proposed for the doubly dynamic traffic assignment to a transportation network are described. In particular, the model has been specified and used to simulate the effects of different control measures, ranging from flexible working times to a trip planning informative system on a small but realistic network.

Although the results are by no means conclusive as a result of the exogenous assumptions made especially about users' behavior, they appear to give some insights about both the potential of the model and the effectiveness of alternative control measures.

The results show that the proposed model is a valid tool to simulate the relevant effects of control strategies in different scenarios. It also appears that some control measures cannot be correctly assessed without the explicit simulation of the demand elasticity over departure times and of the day-to-day adjustment process based on users' memory and forecasting.

The effectiveness of an informative system appears to be greatly affected by the type of users' behavior and by the informative strategy followed. Also, the type of control strategy and memory depth play an important role on the network performance.

The results suggest that a careful evaluation is needed to assess the effects on the network performance and the benefits of informative control strategies, and that the impacts on both informed and noninformed users should be taken into account.

ACKNOWLEDGMENT

Partially supported by research project, PROMETHEUS, under grants 89.01455.93 and 90.01589.93, CNR-ITALY.

REFERENCES

1. E. Cascetta. Static and Dynamic Models of Stochastic Assignment to Transportation Networks. In *Flow Control of Congested Networks* (G. Szaego, L. Bianco, and A. Odoni, eds.). Springer Verlag, Series F, Vol. 38, 1987.
2. E. Cascetta. A Stochastic Process Approach to the Analysis of Temporal Dynamics in Transportation Networks. *Transportation Research*, Vol. 23B, No. 1, 1989, pp. 1–17.
3. E. Cascetta and G. Cantarella. A Day-To-Day and Within-Day Dynamic Stochastic Assignment Model. *Transportation Research*, forthcoming.
4. J. L. Horowitz. The Stability of Stochastic Equilibrium in a Two-Link Transportation Network. *Transportation Research*, 18B, 1984, pp. 13–28.
5. Y. Sheffi. *Urban Transportation Networks*. Prentice-Hall, Englewood Cliffs, N.J., 1985.
6. R. Hammerslag. A Three-Dimensional Assignment in the Time-Space. UTSG Annual Conference 1988, London, England, 1988.
7. B. N. Janson. *Dynamic Traffic Assignment for Urban Road Networks*. Working Paper. Oak Ridge National Laboratory. Oak Ridge, Tenn, 1989.
8. M. Ben Akiva and A. De Palma. *Dynamic Models of Transportation Networks*. PTRC Transport and Planning Summer Annual Meeting, 1987.
9. P. K. Vythoulkas. (1990). A Dynamic Stochastic Assignment Model for the Analysis of General Networks. *Transportation Research*, Part B, forthcoming.
10. K. Small. The Scheduling of Consumer Activities: Work Trips. *The American Economic Review*, Vol. 72, No. 3, 1982, pp. 467–479.
11. H. S. Mahmassani and G. L. Chang. Experiments With Departure Time Choice Dynamics of Urban Commuters. *Transportation Research*, Vol. 20B, No. 1, 1986, pp. 297–320.
12. H. S. Mahmassani and G. L. Chang. Travel Time Prediction and Departure Time Adjustment Behavior Dynamics in a Congested Traffic System. *Transportation Research*, Vol. 22B, 1988, pp. 217–232.
13. M. Ben Akiva, M. Cyna, and A. De Palma. Dynamic Model of Peak Period Congestion. *Transportation Research*, Vol. 18B, Nos. 4–5, 1984, pp. 339–355.
14. Y. Iida, T. Akiyama, and T. Uchida. *Experimental Analysis of Dynamic Route Choice Behaviour*. Preprints of International Conference on Dynamic Travel Behavior Analysis. Kyoto, Japan, July 1989.
15. G. E. Cantarella, E. Cascetta, and M. Di Gangi. An algorithm for a dynamic traffic assignment model. *Proceedings of the Annual Conference of the Operational Research Society of Italy*, Sorrento, Italy, Oct. 1990.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

Effectiveness of Information Systems in Networks With and Without Congestion

RUDI HAMERSLAG AND ERIC C. VAN BERKUM

The use of road transport informatics (RTI) is a recent development that optimizes the use of existing facilities in the transportation system and serves three main goals: alleviation of congestion, diminution of air pollution, and reduction of incidents. RTI instruments deal with traffic information. Examples of RTI systems are pretrip planning, roadside displays, radio data system–traffic message channel, and in-car navigation. To model the effects of providing the road user with information a method is used in which stochastic and deterministic assignments were compared for both networks with and without congestion. To let information also effect destination choice and the spatial distribution of activities, the assignment models were combined with different distribution models. The amount of information that travelers have was translated to a "level of uncertainty" measure. The more informed a traveler is, the lower the level of uncertainty. Since the effects appeared to be network dependent, a number of different networks were examined. Simulations show that the amount of kilometers driven decreases when travelers are provided with better and more information.

The use of road transport informatics (RTI) is a recent development that optimizes the use of existing facilities in the transportation system and serves three main goals: alleviation of congestion, diminution of air pollution, and reduction of incidents. RTI instruments deal with traffic information. Systems such as pretrip planning, roadside displays, radio data system–traffic message channel, and in-car navigation are all part of RTI. From a planners' viewpoint, it is essential to know the possible impact of RTI on the traffic system. One way to predict effects of RTI is to model individual travel behavior and to incorporate information explicitly as a model component. In this way, the effect of information on travel behavior can be simulated. Before this can be done, however, it is necessary to model the current situation, in which the traveler is not perfectly informed and therefore makes nonoptimal choices.

In many existing models it is assumed that people have perfect knowledge of all travel alternatives. This assumption means that the usefulness of providing information to travelers cannot be determined. In the approach presented in this paper, the classic four-stage model is central. The key issue is, however, that the perceived travel times instead of the objective travel times are being used in all stages. Therefore, a measure of uncertainty is introduced. Uncertainty affects not only route choice, but also destination choice and the spatial distribution of activities. A further assumption is that information reduces uncertainty. So by using models in which route choice, distribution, or location of activities, or all three, are influenced by the (perceived) travel times and the outcomes for different levels of uncertainty are compared, it is possible to get an insight in the effects of information.

R. Hamerslag, Departments of Civil Engineering and of Mathematics and Informatics, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands. E. C. van Berkum, Bureau Goudappel Coffeng, P.O. Box 161, 7400 AD Deventer, The Netherlands.

RELATED STUDIES

In recent years, many approaches have been presented to provide insight into the possible benefits of information systems in transport.

The feasibility of the Comprehensive Automobile Traffic Control project (*1*) was studied by using a simulation model in which the noninformed users choose their route on the basis of various factors, such as travel time, length of the route, number of lanes, number of turns, and so on, and the informed users choose their route solely on the basis of travel time. It was found that in Tokyo travel time could be reduced by 6 percent and fuel consumption by 5 percent. Tsuji et al. (*2*) investigated the effectiveness of a route guidance system by using a mathematical model. Among other factors, they used travel time reduction as a measure of effectiveness. The outcomes, however, must be related to the heavy assumptions under which the model is valid. The reduction in travel time was found to be 11 percent. van Vuren (*3*) tried to model the effectiveness of route guidance by using a multiuser class equilibrium and stating that the noninformed users behave greedily, as in an deterministic user equilibrium, whereas the informed users behave according to the principle of a system optimum. The results were found to be unrealistic because the uninformed users were better off.

Koutsopoulos and Lotan (*4*) modeled the impact of information on travelers by using a stochastic user equilibrium and stating that information systems reduce the variance in travel time. They found a reduction in travel time of about 5 percent, dependent on the assumed reduction in variance.

Mahmassani and Jayakrishan (*5*) modeled the effectiveness of a real-time information system on a small test network with three parallel highways and a number of switching possibilities. The researchers chose one information supply strategy and focused on the users' reaction by defining them as bounded rational individuals. An important result was that the system performance might actually worsen by myopic local actions of the drivers. Van Berkum and van der Mede (*6*) presented a dynamic approach that simulates rational, uncertain, persistent individuals who base their decisions on experience and have a limited knowledge of alternatives.

The approach that is followed in this paper is an extension of the work of Koutsopoulos and Lotan (*4*). The situation of recurrent congestion was also studied in this research. But whereas Koutsopoulos and Lotan restricted the effects of better information to route choice, the impact on destination choice and the location of activities has also been studied here. Another difference is that the amount of uncertainty in their approach was initially too small. Further, they examined one network, whereas different networks are studied here. Because the results are network dependent, it is difficult to compare results, but the results they found on route choice are on the same order of magnitude as the results presented in this paper. The results gained from the present research are not comparable with the results found by Mahmassani and Jayakrishan. They studied the reaction of people on dynamic traffic information that reports the actual traffic conditions. When drivers react myopically, this information becomes invalid. An adjustment process will occur, which in the end will lead to an equilibrium. This equilibrium is focused on in this paper. This further implies that the information given to the drivers is in some sense not real-time information but rather future-time information.

MODELING APPROACH

General

The main hypothesis of this study is that the fact that people are uncertain about travel times on links has more effects than only on route choice. There will also be effects on destination choice as well as on the spatial distribution of activities. People make trips because they want to perform activities that are spatially separated. In the traditional four-stage models, the spatial distribution of activities is fixed. In this study models are used that include the spatial distribution of activities as endogenous. Users choose a route by minimizing some measure of cost. In this study travel time will only be used as cost. Travelers do not possess perfect information about the network they travel on. This means that people do not minimize the objective time but rather the perceived travel time.

Destination choice can also be modeled by using a cost minimization procedure (*8*). Because of the observation made previously, this means that in determining the origin-destination (O-D) flows, the perceived cost or travel time must also be used. A basic assumption here is that route choice is made on the basis of the same perceived travel times as destination choice and the location of activities are made. Traffic information affects the perception of travel times in the network. The perceived travel times will be modeled as stochastic variables whose distribution is influenced by the amount of available information.

The approach that has been followed uses the traditional four-stage model as a basis, although an adjusted form has been developed. The following assumptions are therefore needed:

- All people base their decisions on what they know; and
- People base their route and destination decisions on the same perceived travel time.

In order to make the approach not too complex the following limitations have been adopted:

- The total number of trips is constant under all levels of uncertainty;
- All people have access to the same level of information;
- Information is assumed to be good and true; and
- No distinction has been made between different modes and purposes.

Route Choice

Link travel times on the network are defined as stochastic variables. The variance in travel times—that travel times are unpredictable to a certain extent—may be understood as uncertainty of travelers. Consequently, users will have different perceptions of travel times on the links.

A deterministic user equilibrium (DUE) can be defined as the situation in which no traveler can improve his or her travel time by unilaterally changing route (*9*). This definition assumes that every traveler has an exact knowledge of travel times and flows on all links in the network. A stochastic user equilibrium (SUE) can be defined as the situation in which every traveler *thinks* that he or she cannot improve the travel time by unilaterally changing routes (*9,10*). This definition assumes that travelers have different perceptions of travel times. Comparing a SUE with a DUE enables estimating the effect of providing information to travelers (or reducing their uncertainty) on the traffic system (*4,11*), because this comparison can be translated as comparing travelers with exact knowledge of all travel times in the network with travelers with different perceptions of travel times in the network.

In networks without congestion the DUE assignment becomes a simple all-or-nothing assignment, where the SUE assignment becomes a classic stochastic assignment (*12,13*). The impedance Z_{ap} of a link a in a network for person p is a function of a number of variables X_{ak} such as time, cost, and distance and their relative importance β_k plus some measure of uncertainty. We define

$$Z_{ap} = \sum_k \beta_k \cdot X_{ak} + e_{ap} \qquad (1)$$

where e_{ap} is a noise term. The resulting route choice model depends on the distribution of e_{ap}. It is supposed that e_{ap} is normally distributed with mean 0 (*13*), which yields a probit model for route choice. The introduction of the noise term e_{ap} can be explained by stating that (*a*) behavior cannot completely be explained by all X_{ak}'s, (*b*) individuals have different perceptions of the X_{ak}'s and their relative importance therefore may differ, and (*c*) individuals are uncertain about the exact value of the X_{ak}'s, especially because these values differ in time. Instead of impedance, generalized cost, or generalized time only travel time will be considered as a measure for deterrence in this study.

The travel time on a link a in a network *without congestion* is

$$Z''_a = Z_a + \alpha \cdot R \cdot \sqrt{Z_a} \qquad (2)$$

where

Z_a = mean travel time of link a,
R = draw from a normal [N(0,1)] distribution, and
α = factor determining the variance (from now α will be called *level of uncertainty*).

The value of α is dependent on the chosen dimension (*14*). Given an O-D matrix, α can be determined by comparing true with model flows. When the dimension is minutes, it has been estimated that $0.5 < \alpha < 1$ for a regional network with relatively few alternative routes (*15*). Furthermore, Bovy (*14*) developed an efficient methodology for estimating α from observed flows.

In reality the uncertainty will, among other things, be a function of the frequency with which a person travels between a certain O-D pair. The lower the frequency, the higher the uncertainty. In this study, the uncertainty is assumed to be equal for all travelers.

The travel time of a link in a network with congestion is

$$Z''_a = Z'_a + \alpha \cdot R \cdot \sqrt{Z'_a} \tag{3}$$

with

$$Z'_a = Z_a\left[1 + \tau\left(\frac{q_a}{c_a}\right)^4\right] \tag{4}$$

where

Z_a = the mean travel time of link a,
q_a = the flow on link a,
c_a = the capacity of link a,
R = a draw from a normal [N(0,1)] distribution,
α = the level of uncertainty, and
τ = a parameter dependent on the definition of capacity.

Destination Choice and the Location of Activities

Because the distribution process is a utility maximization process (or disutility minimization), information will also have impact on destination choice resulting in a distribution of flows and the location of activities. In this study, the following interaction model with elastic constraints is used (*16*):

$$\min \sum_j \left(\sum_i T_{ij} - m_j^{-h} \cdot A_j\right)^2 + \sum_i \left(\sum_j T_{ij} - l_i^{-g} \cdot D_i\right)^2 \tag{5}$$

Subject to

$$T_{ij} = \theta l_i m_j Q_i X_j \exp[-0.4 \ln^2(Z_{ij} - d_{ij} + 1)] \tag{6}$$

where

T_{ij} = number of trips between i and j,
l_i, m_j = equilibrium factors,
Q_i, X_j = polarities,
Z_{ij} = objective travel time between i and j,
A_j, d_i = arrivals and departures, and
d_{ij} = difference between objective and perceived travel time between zones i and j.

In solving the model, the terms $1/(1 + g)$ and $1/(1 + h)$ become important. These terms will be called elasticities. Thus when g and h are both 0 the elasticities become 1 and the model turns into the classic gravity model with fixed constraints. To coordinate spatial planning, transportation development, and spatial development, the model with elastic constraints was developed. The value of the equilibrium factors in Equation 6 is a function of the extra effort needed to comply with the constraints. In poorly accessible areas, the value is high and, inversely, in easily accessible areas the value is low. When the number of arrivals and departures is seen as dependent, though not exclusively, on the accessibility, the objectives in Equation 5 must become elastic.

Combining the Assignment and the Distribution Model

To determine the effects of information on route choice, route and/or destination choice, and/or the location of activities, the following models must be compared:

- In the case of *no* congestion, a distribution model with and without elastic constraints will be compared with the same model but combined with a stochastic Burrel assignment.
- In the case of congestion, first the DUE assignment will be combined with the distribution model without (*10,17*) and with elastic constraints (*18*). Second, the same combination will be made, but with the SUE assignment.

To combine a SUE assignment with a distribution model, including the assumption that both models deal with the same perceived travel times, it is necessary to determine how the perceived travel times must be used in the distribution stage. In the proposed distribution model there is one value for travel time between each O-D pair. In reality this travel time is different for every individual (perceived travel time). Starting with Z_{ijrp}, the perceived travel time between i and j along route r of person p, person p chooses that route with the smallest perceived travel time. Therefore, it holds that

$$Z_{ij,p} = \min_r Z_{ijrp} \tag{7}$$

Suppose the population B is divided in two groups, B1 and B2. Persons belonging to B1 find route *1* the best, and persons belonging to B2 do not, so

$$Z_{ij1p} \leq Z_{ijrp} \;\forall\, p \in \text{B1 and } r \neq 1 \tag{8}$$

For persons belonging to B2 it holds that

$$Z_{ij1p} \geq Z_{ij,p} \;\forall\, p \in \text{B2} \tag{9}$$

So

$$Z_{ij,p} \leq Z_{ij1p} \;\forall\, p \in (\text{B1} \cup \text{B2}) \tag{10}$$

Suppose there are N persons in B, then

$$Z_{ij} = \frac{1}{N}\sum_p Z_{ij,p} \text{ and } Z_{ij1} = \frac{1}{N}\sum_p Z_{ij1p} \tag{10.1}$$

Using Equation 10 it holds that

$$Z_{ij} \leq Z_{ij1} \tag{10.2}$$

The same result can be derived for every route r, so

$$Z_{ij} \leq Z_{ijr} \qquad \forall\ r \tag{11}$$

Thus the perceived travel time between any O-D pair used in the distribution stage is always less than or equal to the perceived travel time of any of the chosen routes between the OD pair.

The difference between the best route and the travel time between an O-D pair is dependent on the network. When, for instance, one route is by far the best so that every traveler between that OD pair will choose that route, the equal sign in Equation 11 holds for this particular route. When there is a spreading over the routes for all r the less than sign will hold. When the level of uncertainty α becomes larger, the spreading in routes becomes larger and Z_{ij} will decrease, or in other words the difference between model travel time and the mean perceived travel time of the objectively seen best route (which is by definition the objective travel time of the best route) becomes larger. So in the distribution stage the following travel time is used:

$$Z_{ij} - d_{ij}$$

Where Z_{ij} is the mean perceived travel time of the objectively seen best route between zone i and j; d_{ij} is an increasing function of α (obviously when $\alpha = 0$, also $d_{ij} = 0$).

Models and Algorithms

To study the effects of more or better information on route choice the "A model" is used, which is a stochastic equilibrium assignment with a given, fixed O-D matrix. To study the effects on destination choice and on the resulting O-D flows too, the "A + D model" is used. In this model, a stochastic equilibrium assignment and distribution with fixed constraints are combined. In the O-D matrix, the numbers of departures and arrivals are fixed for each zone. The cell volumes solve Equation 5 subject to Equation 6 with $g = h = 0$.

To study the effects of activities on the locations, the "$A + D + L$ model" is used. In this model, a stochastic equilibrium assignment and distribution with elastic constraints are combined. In the O-D matrix, the numbers of departures and arrivals are variable for each zone, but the total number of trips is fixed. The cell volumes solve Equation 5 subject to Equation 6 with g and h not necessarily equal 0.

In Figures 1 through 4, the separate algorithms for the congestion situation are depicted. Basically, the methodology as proposed by Evans (*17*) is followed. The steps that have to be executed more than once because the draw must take place m times have been depicted with a thick line. In the case of no congestion, the step where new travel times are computed becomes trivial.

A generalized description of the used algorithm is

1. Read network;
2. Draw link travel times for every link;
3. Determine travel times from shortest routes between every O-D pair;
4. Repeat No. 2 and No. 3 m times;
5. IF model = A THEN
   read O-D matrix
   ELSE determine mean travel times with the travel times per draw determined in No. 3. Determine O-D matrix with elastic constraints (A + D + L) or with fixed constraints (A + D) using a Gauss-Seidel interation procedure to solve Equation 5 subject to Equation 6;

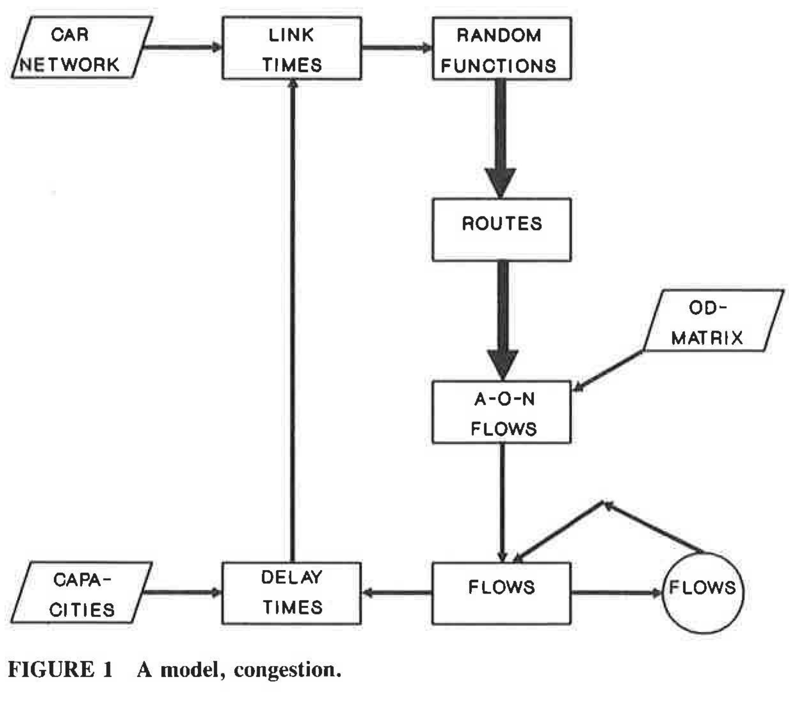


FIGURE 1 A model, congestion.

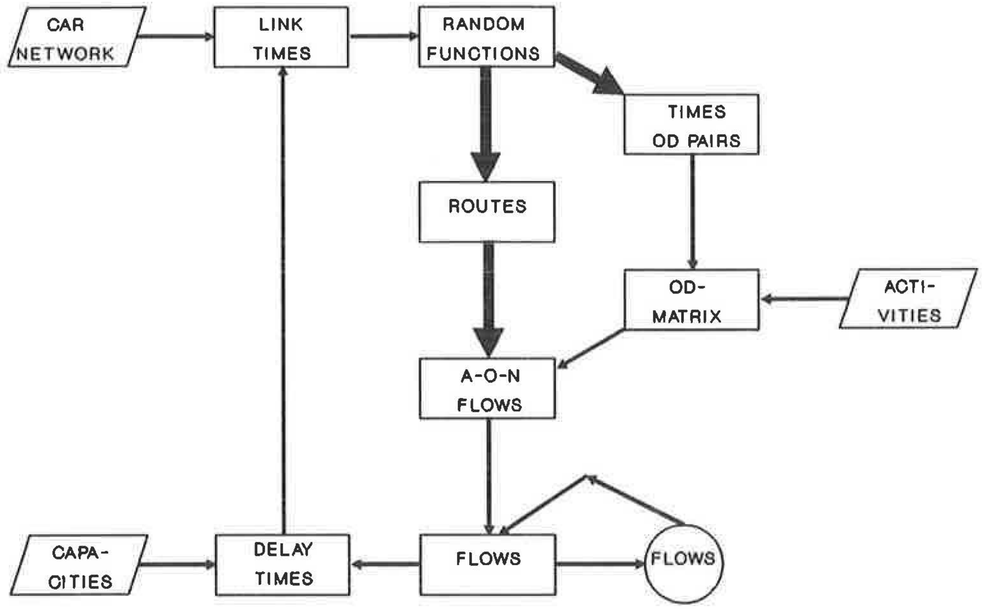


FIGURE 2 A + D model, congestion.

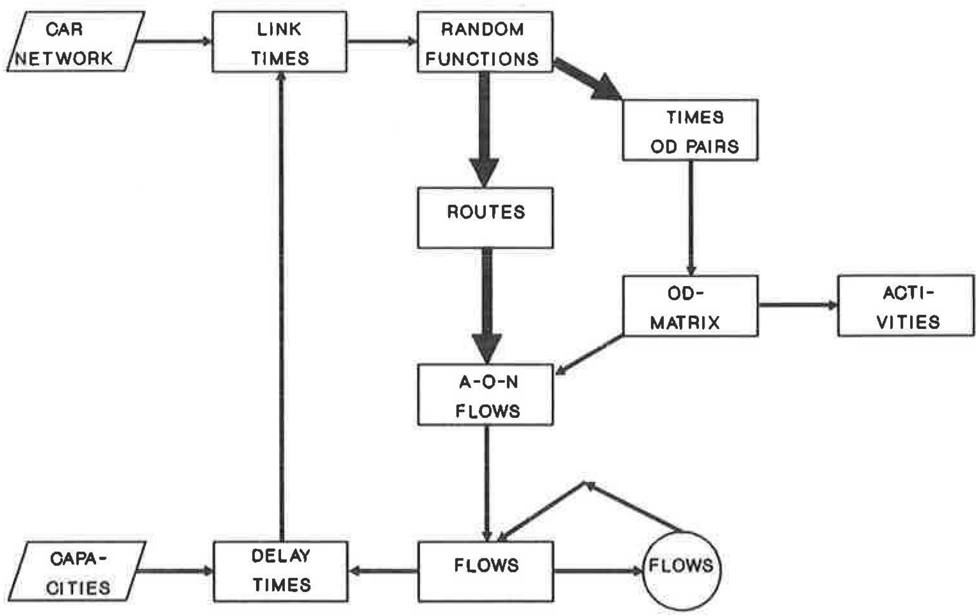


FIGURE 3 A + D + L model, congestion.

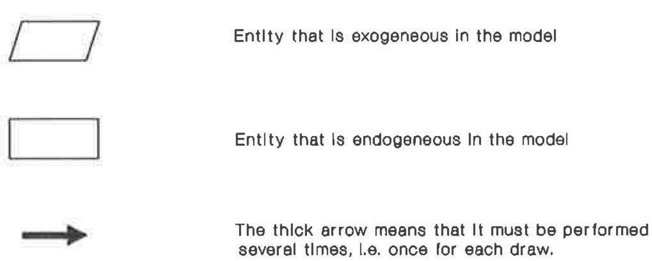


FIGURE 4 Explanation for Figures 1 to 3.

6. Subdivide the O-D matrix in m equal parts and load them to the routes determined in No. 3, yielding loads q_a^+ for link a;
7. load link a in iteration i the network with $q_a^i = [q_a^{i-1} \cdot (i - 1) + q_a^+]/i$;
8. In case of congested networks: determine new travel times; and
9. Go to No. 2 until stop criterion is reached.

EXPERIMENTS

The experiments were performed using the research facilities of the Teacher Friendly Transportation Programs V90.2 (*19*). In the stochastic assignments m, the number of draws was 4 and the number of iterations was 8. Because for every tree of shortest paths, new travel times were drawn, the total number of draws is 32 times the number of zones. Convergence was no problem in all test networks. The number of iterations was far less than expected in a combined distribution-assignment procedure (*10,20*).

Networks

Earlier it was observed that the spreading of chosen routes determines to some extent the value of d_{ij}. The amount of spread is not only dependent on the size of the variance as used in the stochastic assignments, but also on the presence of (relevant) alternative routes. Obviously in a situation in which there are hardly any alternative routes, the spread will be small. Therefore it is important to investigate different networks. In this study four regional networks with a diameter of about 40 km (called REGIO, RING, SLOW, and CBD) and one urban network with a diameter of about 15 km (TOWN) were examined. For the regional networks only, the situation without congestion is considered. For the urban network both the situation with and without congestion are considered.

In Figures 5 to 7 some of the networks with their spreading of activities and flows are shown. In Figure 8 the notation of the activities is shown. The networks RING, SLOW, and FAST are the same size (number of links, number of nodes, distances) as CBD.

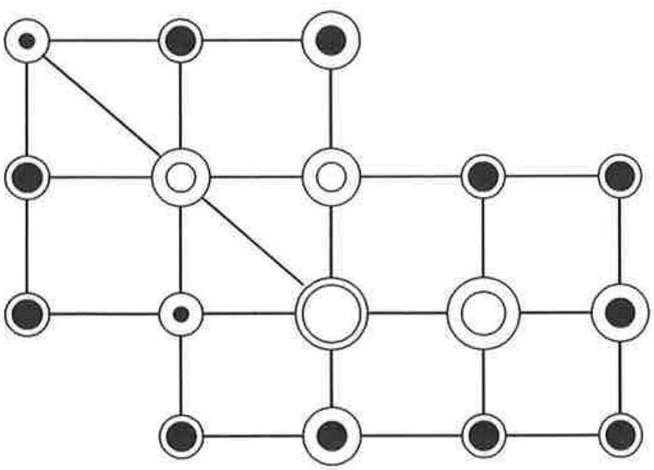


FIGURE 5 Network CBD.

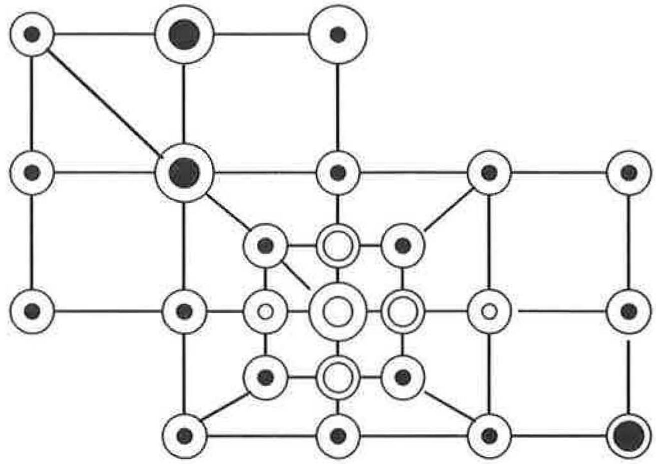


FIGURE 6 Network REGIO.

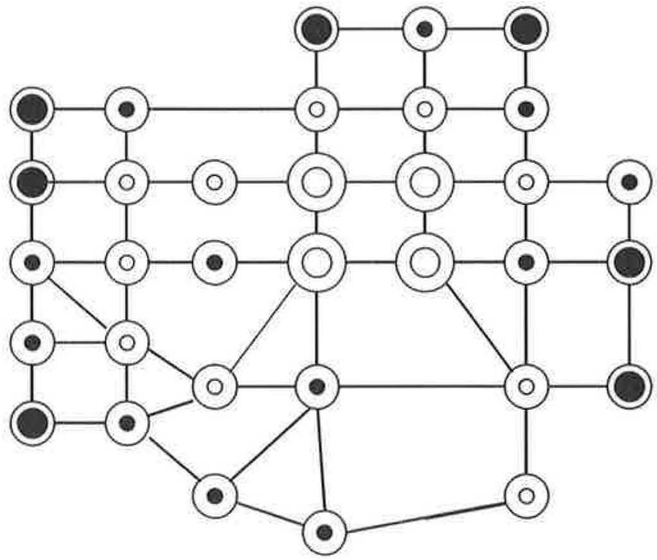


FIGURE 7 Network TOWN.

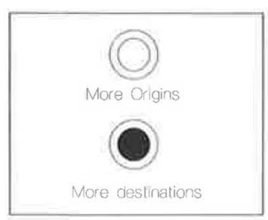


FIGURE 8 Notation of activities.

In Figure 8 the radius of the outer circle is proportional to

$$\left[\max\left(\sum_i T_{ij}, \sum_j T_{ij}\right)\right]^{1/2} \tag{12}$$

The radius of the inner circle is proportional to

$$\left[abs\left(\sum_i T_{ij} - \sum_j T_{ij}\right)\right]^{1/2} \tag{13}$$

The inner circle is open when

$$\sum_i T_{ij} < \sum_j T_{ij} \tag{14}$$

The description of the networks is as follows:

- CBD—network with speedways (100 km/h) pointed to a central zone; other roads are 40 km/h;
- FAST—network with only speedways (100 km/h);
- REGIO—network like "CBD", but more dense near the center;
- RING—network with speedways in a ring around a central area;
- SLOW—network with only secondary roads of 40 km/h; and
- TOWN—urban network; there are no trips to or from the surrounding areas.

Level of Uncertainty, α

The simulations have been performed for all networks with levels of uncertainty $\alpha = 0$, $\alpha = 0.5$, and $\alpha = 1$. For the network TOWN $\alpha = 0.3$ and $\alpha = 0.8$ were also taken into account.

Results

The results of the simulations for the urban network with and without congestion are given in Tables 1 and 2. Results for the regional networks without congestion are given in Tables 3, 4, and 5. For all networks, the amount of carkilometers (vehicle miles) increases when the level of uncertainty increases. Because provision of information can be translated in a smaller level of uncertainty, it can be stated that providing road users with information reduces the amount of carkilometers. The results show that the gains differ per network. A network means not only the set of links and nodes, but also the initial tripends. This observation implies that it is hard to compare the results of other studies with one another and with these results, because different networks are used in all studies.

The results for the networks as listed in Tables 3, 4, and 5 are more or less comparable. These results were calculated with models that did not deal with congestion. The results for the TOWN network show a larger increase in carkilometers when α increases (See Table 2). This can be explained by the fact that the TOWN network obviously contains more alternative routes than all the other networks. The spread in route choice will be bigger for this network since there simply exist more alternatives. Because the network outcomes in Tables 3 to 5 reflect few alternative routes, the effect on route choice is small compared with the effect on destination choice (compare the outcomes in Tables 3 and 4). The extra effect on the location of activities is also small compared with the effect on destination choice (compare the outcomes in Tables 4 and 5). When looking at the network TOWN, the effects on route choice are the largest. Change in destination choice and in the location of activities are marginal compared with this effect. Because this network is more realistic than the other ones, this observation may be generally true. By comparing Tables 1 and 2, it follows that the effect of the provision of information is larger in the network with congestion than without congestion.

TABLE 1 CARKILOMETERS FOR TOWN NETWORK, WITHOUT CONGESTION, UNDER DIFFERENT LEVELS OF UNCERTAINTY (KM FOR $\alpha = 0$ ARE 100)

α	A	A+D	A+D+L
1.0	117	128	130
0.7	112	119	121
0.5	104	107	108
0.3	101	102	103
0.0	100	100	100

TABLE 2 CARKILOMETERS FOR TOWN NETWORK, WITH CONGESTION, UNDER DIFFERENT LEVELS OF UNCERTAINTY (KM FOR $\alpha = 0$ ARE 100)

α	A	A+D	A+D+L
1.0	124	129	136
0.7	118	121	126
0.5	106	109	110
0.3	102	103	103
0.0	100	100	100

TABLE 3 CARKILOMETERS DRIVEN FOR DIFFERENT NETWORKS, UNDER DIFFERENT LEVELS OF UNCERTAINTY WITH THE A-MODEL (KM FOR $\alpha = 0$ ARE 100)

α	CBD	RING	SLOW	FAST	REGIO
1.0	102	103	101	105	103
0.5	100	103	101	101	101
0.0	100	100	100	100	100

TABLE 4 CARKILOMETERS DRIVEN FOR DIFFERENT NETWORKS, UNDER DIFFERENT LEVELS OF UNCERTAINTY WITH THE A+D-MODEL (KM FOR $\alpha = 0$ ARE 100)

α	CBD	RING	SLOW	FAST	REGIO
1.0	108	109	112	120	111
0.5	102	103	105	107	103
0.0	100	100	100	100	100

TABLE 5 CARKILOMETERS DRIVEN FOR DIFFERENT NETWORKS, UNDER DIFFERENT LEVELS OF UNCERTAINTY WITH THE A+D+L-MODEL (KM FOR $\alpha = 0$ ARE 100)

α	CBD	RING	SLOW	FAST	REGIO
1.0	113	113	113	123	115
0.5	103	103	105	107	104
0.0	100	100	100	100	100

CONCLUSION

Trips and activities are a result of decisions people make. These decisions concern route and destination choice as well as activity choice. The actual choices depend on the perceived travel times, rather than on the objective travel times. As a result, travelers think they choose the best route, but this route is not necessarily the best from an objective point of view. Also destinations are chosen because they appear to be close. This causes extra, unnecessary carkilometers.

The approach presented in this paper has a number of assumptions and limitations about information:

- Information is seen as an abstract entity; it is not possible to evaluate a specific information system or different types of information.
- Because of the equilibrium approach the presented method is able to predict the long-term effects of the provision of information in a situation of recurrent congestion.

The results of this study should be looked at in light of these assumptions as well as in light of the limitations this approach has.

It was proven that the perception of two or more independent routes is always less than or equal to the perception of each of two or more routes together. The travel time of the chosen route is systemically being underestimated. Providing information reduces the difference between perceived travel time and objective travel time. This has an impact on the choice of route, destination, and activity. As a result, the amount of carkilometers decreases. The different test cases show that the form of the network, with respect to the presence of alternative routes, is of importance. Further, the simulations show that in a situation with congestion, the decrease of carkilometers is larger than in the situation with no congestion. Currently it is not possible to quantify the effects of information precisely because the present and future values of α are not exactly known, uncertainty will only partially be influenced by information, and only a part of the travelers will use the information. On the other hand, through route guidance, delays on intersections may be minimized (*21*) and the influence that information about incidents could have is neglected. With the above considerations in mind it seems valid to state that information systems may decrease the amount of carkilometers in urban networks by 15 to 20 percent and in regional networks by 5 to 10 percent.

REFERENCES

1. F. Kobayashi. Feasibility Study of Route Guidance. In *Transportation Research Record* 737, TRB, National Research Council, Washington, D.C., 1979, pp. 107–112.
2. H. Tsuji, R. Takahashi, H. Kawashima, and Y. Yamamoto. A Stochastic Approach for Estimating the Effectiveness of a Route Guidance System and its Related Parameters. *Transportation Science*, Vol. 19, No. 4, 1985, pp. 333–351.
3. T. van Vuren. Combined Assignment (in Dutch). In *Colloquium Vervoers-planologisch Speurwerk* (P.M. Blok ed.), Delft, The Netherlands, 1988, pp. 731–746.
4. H. N. Koutsopoulos and T. Lotan. Effectiveness of Motorist Information Systems in Reducing Traffic Congestion. Presented at IEEE Vehicle Navigation and Information Systems Conference (VNIS '89), Toronto, Canada, 1989.
5. H. S. Mahmassani and R. Jayakrishan. Dynamic Analysis of Traffic Network Performance under Real-Time Information. Paper presented at TRB 69th Annual Meeting of the Transportation Research Board, Washington, D.C. 1990.
6. E. C. van Berkum and P. H. J. van der Mede. A Dynamic Behavioral Model for Route, Mode and Departure Time Choice. Presented at PTRC Conference, Brighton, England, 1990.
7. EUROTOPP. *Interactions with RTI-Technologies*, Report to DRIVE, Bureau Goudappel Coffeng, Deventer, 1990.
8. A. G. Wilson. *Urban and Regional Models in Geography and Planning*. Wiley and Sons, London, England, 1974.
9. Y. Sheffi. *Urban Transportation Networks*. Prentice-Hall, Englewood Cliffs, N.J. 1985.
10. A. J. Horowitz. The Stability of Stochastic User Equilibrium in a Two-Link Network, *Transportation Research*, Vol. 18B, 1984, pp. 13–28.
11. Y. Sheffi and W. B. Powell. A Comparison of Stochastic and Deterministic Traffic Assignment over Congested Networks. *Transportation Research*, Vol. 15B, 1981, pp. 53–64.
12. J. E. Burrel. Multiple-Route-Assignment: a Comparison of Two Methods. In *Traffic Equilibrium Models* (M. Florian ed.), Springer Verlag, Berlin, 1974, pp. 229–239.
13. H. von Falkenhausen. *Ein Stochastisches Modell zur Verkehrsumlegung*, Strassenbau und Strassenverkehrtechnik, Koeln, Heft 58, 1967.
14. P. H. L. Bovy. *Traffic Assignment in Uncongested Networks*. Dissertation. Delft University of Technology, 1990.
15. R. Hamerslag. Investigation into Factors Affecting the Route Choice in "Rijn-streek-West" with the Aid of a Disaggregate Logit Model. *Transportation*, Vol. 10, 1981, pp. 373–391.
16. R. Hamerslag. Spatial Development, Developments in Traffic and Transportation, and Changes in the Transportation System. In *Changes in the Field of Transport Studies* (J. B. Polak and J. B. van de Kamp eds.), Nijhoff, The Hague, 1980, pp. 60–80.
17. S. P. Evans. Derivation and Analysis of Some Models Combined Trip Distribution and Assignment. *Transportation Research*, Vol. 10, 1976, pp. 37–57.
18. E. C. van Berkum, H. J. Meurs, and R. Hamerslag. Dynamics in Transport Models. In *Colloquium Vervoers-planologisch Speurwerk* (H.J. Meurs, ed.) Delft, The Netherlands, 1989, pp. 413–424.
19. R. Hamerslag. Teacher Friendly Transportation Programs, *Microcomputers in Civil Engineering*, Vol. 3, 1988, pp. 81–89.
20. A. J. Horowitz. Convergence Properties of some Iterative Traffic Assignment Algorithms. In *Transportation Research Record 1220*, National Research Council, Washington DC, 1989.
21. R. Hamerslag. Controlled Traffic with Information Systems and Model Applications (in Dutch). In *Symposium Beheers(t) Verkeer* (F. Middelham, ed.), Delft University, Delft, 1990.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

Multiple User Class Assignment Model for Route Guidance

Tom van Vuren and David Watling

The application of the concept of multiple user class equilibrium assignment to the modeling of route guidance systems is the concern of this paper. In particular, its role in modeling guided and unguided drivers is discussed, as well as its ability to lead to guidance strategies that are effective even with a large proportion of drivers equipped. A review of previous route guidance model work is given. A multiple user class model of route guidance is then proposed and the properties of such a model are discussed. Finally, a presentation of simulation results obtained from such a model, for two real-life networks and for a number of route guidance scenarios, is given.

One of the basic implicit assumptions in standard assignment methods is that drivers and vehicle attributes are identical; they do not differ from one another in either their travel cost definition or their vehicle size or vehicle performance.

APPLICATION OF MULTIPLE USER CLASS ASSIGNMENT TO ROUTE GUIDANCE

Dafermos (*1*) was probably the first to realize the limitations of this assumption, and to propose as a remedy a multiple user class (MUC) model, which takes differences between drivers and between vehicles into account. These classes may differ in (*a*) vehicle type or size, (*b*) travel cost definition, and (*c*) network restrictions (*2*). Within each class, however, driver and vehicle attributes are identical. Typical classes could be lorries (particularly in conjunction with lorry bans), commuters (minimizing some measure of generalized cost), business travelers (minimizing travel time), and tourists (following road signs).

In a MUC assignment model all classes are to be assigned to the network in interaction with each other, so that in equilibrium for each class "no-one can improve his or her (perceived) travel cost by unilaterally changing route," and in that respect MUC assignment is clearly an extension to the standard, single class assignment model.

The relevance of the MUC concept for route guidance modeling is evident. In a situation with some kind of in-vehicle route guidance at least two user classes can be defined: those who are equipped, and those who are not. In fact three groups could even be distinguished; namely, those who follow complete guidance, those who follow partial guidance (because they either lose their way or their confidence in the advice), and those who do not follow guidance at all. Each of these user classes would have a different cost definition and possibly even different network restrictions (e.g., if the guidance network does not include all existing roads for environmental or computational reasons).

A number of cost definitions for each user class can be distinguished on the basis of one's representation of route choice in reality and the assumed routing criterion in the guidance system. The assumptions made in our model are illustrated in the following table and will be discussed next.

	Case A	*Case B*	*Case C*
Guided	UE	SO	SUE
Unguided	SUE	SUE	SUE

The authors believe that drivers currently make perceptual errors in their routing decisions. If it is assumed that these will be removed by the guidance system, a combination of stochastic user equilibrium (SUE) and user equilibrium (UE) assignment, as in case A, will occur. Along the same lines, a combination of guided drivers following a system optimum could be envisaged, so as to improve network conditions explicitly, and unguided ones following a SUE route pattern, as in case B.

Then, if it is expected that the system not provide flawless information (because of communication delays or forecasting errors), the two classes could both follow a SUE (case C), but the variance in randomized link travel times would be lower for the equipped vehicles. In fact, the model presented here allows for all three guidance criteria to be implemented simultaneously, to given proportions of the equipped drivers, although for the purposes of the numerical results described later, the three cases will only be examined individually. For brevity, throughout this paper cases A, B, and C will often be referred to as the UE routing strategy, the SO routing strategy, and the SUE routing strategy, respectively.

The tool of MUC assignment now allows the investigation of the various future scenarios by comparing resulting network costs of different guidance strategies, and by changing the proportion of drivers per user class, the spread in link time errors, and the level of congestion, using real-life networks. In this paper, the development of such a route guidance simulation model, based on MUC assignment, will be described and selected simulation results will be presented.

PREVIOUS ROUTE GUIDANCE MODEL WORK

Earliest reported route guidance related model work was carried out by Kobayashi (*3*) who used a simulation model to

Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, England. T. van Vuren's current affiliation: Frank Graham Consulting Engineers, Elgar House, Shrub Hill, Worcester WR4 9EN, England.

assess benefits of the Comprehensive Automobile traffic Control System (CACS) route guidance system by comparing shortest routes for guided drivers with routes for nonguided drivers on the basis of road length, number of lanes, percentage of trunk lanes, and number of right and left turns. The network consisted of 99 intersections and 286 directional links; Kobayashi estimated a maximum total possible reduction in travel time in this network of 6 percent, at a level of take-up of 50 to 75 percent, compared with observed travel times in Tokyo. Tsuji et al. (*4*) set up a mathematical model, on the basis of the stochastic nature of travel time, even under guidance. The proportion of guided vehicles was assumed so small, that no influence on nonguided vehicles was expected. Comparing the expected travel time of guided routes with those of alternative routes, they estimated an expected reduction in travel time for guided drivers of approximately 11 percent, which compared well with the observed reduction in the CACS system of some 12 percent. Al-Deek et al. (*5*) compared shortest route travel times with the observed route pattern using TRANSYT-7F in the Los Angeles SMART corridor, and found that for recurring congestion the travel time savings of route guidance would be negligible (less than 3 min per trip of on average 25-min length). A similar approach was adopted by Rakha et al. (*6*) who compared routes based on free flow costs (for unguided drivers) and those based on minimum costs (for guided drivers). These assumptions are clearly not valid in congested situations, which probably accounts for the possible total network travel time savings they recorded of up to 21 percent. An interesting finding of their simulations was, however, that a large proportion (85 percent) of total possible savings was achieved with the first 20 percent of equipped vehicles.

Breheret et al. (*7*) used the heuristic dynamic assignment model CONTRAM (*8*). They assumed unguided drivers to follow an approximate stochastic user equilibrium on the basis of prevailing conditions, whereas guided drivers followed optimum routes on the basis of current conditions. If multiple routes were calculated in an attempt to find a user equilibrium reassignment, guided drivers obtained travel time benefits of up to 15 percent; most of these benefits were obtained before a level of take up of guidance of 10 percent of the driver population. In this model, the reassignment of guided drivers benefited nonequipped drivers also with travel time reductions of up to approximately 4 percent. Finally, under these assumptions most of the total network travel time savings (of up to some 5 percent of total travel time) was achieved at a level of take-up of approximately 20 percent [see for e.g., Rakha et al. (*6*)]. If, however, reassignment for guided drivers was calculated by a single, shortest route, total network, travel times invariably increased, which indicated possible problems with systems that advise single routes. Even when unguided drivers were allowed to reassign because of the changed conditions, resulting network travel time savings of such a system were negligible or negative. Smith and Russam (*9*) also reported on a CONTRAM-based model study, in their case of the possible benefits of AUTOGUIDE in London. Whereas unguided vehicles were assumed to base their routes on the average demand pattern and subsequent link costs, guided vehicles were routed along actual optimum routes (for a randomly perturbed trip matrix). They found an estimated average journey time saving of 6 to 7 percent for guided vehicles, which actually decreased with an increase in take-up. Unequipped vehicles benefited also by the guidance system, with travel time reductions of up to 3 percent, resulting in overall network travel time savings of 2.5 to 6.0 percent.

Van Vuren et al. (*10*) employed a MUC assignment model to investigate the situation in which unguided drivers follow a user equilibrium while guided drivers are assigned according to a system optimum. These researchers were mainly concerned with conditions on the cost function that guaranteed existence and uniqueness of a solution. Their main finding was that for a combination of system optimal and user equilibrium drivers and for the family of cost functions considered, only a polynomial cost function gave rise to a convex minimization problem (hence guaranteeing existence and uniqueness properties). Further discussion of these findings appears subsequently. Numerical results given by Van Vuren et al. were limited. Koutsopoulos and Lotan (*11*) assumed that route guidance would reduce the perception errors in link travel time estimates by participating drivers, so that their model consists of a stochastic user equilibrium assignment of two user classes with different variances in the (normal) distribution of random perturbations in perceived link costs; these cost functions did not satisfy conditions for existence and uniqueness, as derived by Daganzo (*12*). Scenarios they investigated on a 204-node network were level of information (influencing the perception errors by guided drivers), percentage of take-up, and the level of recurring congestion. Clearly, an increase in the quality of information resulted in a reduction in perception errors by guided drivers, and therefore in a reduction in their travel times. The advantage of guided drivers over unguided drivers in average travel time was roughly 4 percent, independent of the level of take-up, and an increase in congestion actually reduced the benefits of route guidance. Although these results are obviously rather limited, the most important finding by Koutsopoulos and Lotan is that in their model unguided drivers did not benefit at all from the improved route choice by the guided vehicles. This finding conflicts with the generally held belief [see for e.g., Jeffery (*13*) and Smith and Russam (*9*)] that route guidance benefits nonusers too.

For a corridor consisting of three parallel highways plus connecting links, Mahmassani and Jayakrishnan (*14*) built a model based on route-switching assumptions for drivers that receive dynamic network information. The main conclusions for this simple network were as follows. For optimum resulting travel times both for the equipped drivers and for the system as a whole, route switching should only take place if the alternative route for the trip remainder is at least 20 percent shorter than the existing route, indicating possible instability problems for systems that advise optimum routes (like ALI-SCOUT and AUTOGUIDE), as compared with systems that provide in-vehicle information (such as PATHFINDER). Second, benefits for individual drivers decreased with an increase in participation, whereas benefits for the system as a whole (generally) increased with such an increase; above 50 percent participation the increase in benefits was negligible.

The results of these various model studies are clearly rather ambiguous. Hypotheses about the route choice and interaction of guided and unguided drivers strongly influence the model outcomes. Often the models used in these studies are heuristic or they are only valid under rather strong assumptions. This is not to belittle the importance of these model studies: it merely shows the current problems in understanding and anticipating the behavior of future route guidance systems.

DEVELOPMENT OF THE MODEL

The MUC model of a route guidance system will now be introduced, building on the concepts described in the first section of this paper. The demand for travel (as represented by the mean origin-destination matrix) is assumed to be fixed, as are the network supply conditions. It is also assumed that the whole network is available to the guidance system. Average cost-flow relationships are supplied for each link. In all cases, "cost" is measured in terms of travel time and so the words cost and time will be used interchangeably.

The model consists of four user classes, the demand level for each being a fixed (known) proportion of the origin-destination matrix. Three of the user classes correspond to vehicles equipped with a guidance device, and for two of these three equipped classes it is assumed further that the guidance system is provided with perfect information and that the guided drivers adhere totally to the route recommendations. The first class consists of unguided drivers, each of which aims to minimize his or her own personal cost of travel, but in general fails to do so because of imperfect knowledge of the traffic conditions. This class is modeled by a SUE, the "perceived cost" for each link following some specified distribution (discussed later). The second class is a subset of the equipped vehicles wherein each driver is guided so as to minimize his own personal travel cost. The perfect information assumed to be available to the guidance system is used to eliminate the perception errors, that is, they follow a Wardrop user equilibrium. The third class consists of a second subset of the equipped drivers, which are guided so as to minimize the total system cost [system optimal (SO)], again using the perfect information available. The fourth class consists of the remaining equipped drivers. The aim of the guidance system for this class is again to recommend routes according to a UE pattern; however, to represent the effect of errors in the journey time prediction methods or of drivers not adhering completely to the recommendations, they are modeled by a SUE, but with a distribution for the stochastic variations, which is different from that of the unguided drivers.

The four user classes interact with one another, in the sense that the flows of one user class affect the costs, and hence the route choice, of the other user classes. In this way, the assumption is that under such steady-state conditions, the unguided drivers will tend to change their routes in response to the new route choice of the guided drivers.

Van Vuren (*10*) concluded that for a guidance system with user equilibrium unguided drivers and system optimal guided drivers, the only link cost functions c_a of the family which was established by Van Vliet et al. (*2*) to ensure existence and uniqueness of a multiple user class equilibrium were of the polynomial form:

$$c_a = d_a + b_a F_a^k \tag{1}$$

where

F_a = the total flow on link a,
d_a = a constant representing fixed effects such as free flow travel time,
b_a = a constant, and
k (>0) = a link independent constant.

In the more general four user class model considered here, the same result of Van Vliet et al. (*2*) cannot be used because the properties were established only for the deterministic cost case, whereas here a mixture of stochastic and deterministic costs are used. Results established by Daganzo (*12*) may be used, however, for a similar (though more general) family of cost functions to those of Van Vliet et al., but for the case in which some of the classes may have stochastic costs. Then, in a similar way to Van Vuren et al., by applying the work of Daganzo, it follows that the equilibrium for our more general four user class model is guaranteed to exist and be unique (with respect to link flows and user class/link costs) for cost functions of the form in Equation 1. In this case, c_a is the actual link travel time for all drivers. In the assignment, however, each class will be associated with a different cost: the unguided drivers making random perception errors with Equation 1 as the mean; the guided SUE drivers experiencing different random errors because of imperfect recommendations, and so on; the guided SO drivers using marginal costs corresponding to the actual costs (Equation 1); and the guided UE drivers using the actual costs (Equation 1). That Daganzo's results may be applied to guarantee the above conditions on the equilibrium may be verified as follows.

It is well known that a system optimal assignment in the one user class case with link costs c_a may be obtained by a user equilibrium assignment with marginal link costs c'_a given by

$$c'_a = c_a + F_a \frac{dc_a}{dF_a}$$

To obtain, then, the required routing pattern with actual link costs (Equation 1), the user class costs c_{ai} for link a and user class i must be

$$c_{a1} = c_{a2} = c_{a4} = d_a + [b_a F_a^k]$$

$$c_{a3} = d_a + (k + 1)\,[b_a F_a^k]$$

where user class 1 consists of the unguided drivers, and the remaining classes are the guided drivers, following (respectively) UE, SO, and SUE routing; perceived costs are therefore stochastic for user classes 1 and 4, and deterministic for user classes 2 and 3. It may be seen that the user class cost functions above are indeed of the form required to apply Daganzo's work. Furthermore, Daganzo's conditions require that the variance of the perceived journey time distribution is flow independent. This condition has been noted variously by authors investigating the single user class stochastic user equilibrium case: Sheffi and Powell (*15*), Daganzo (*16*) and Sheffi (*17*). In the last reference, Sheffi suggests—for a probit-based route choice model—the use of a standard deviation of link a perceived cost of θc_{oa}, where c_{oa} is the free-flow travel time and $\theta(>0)$ is a constant. In the guidance model proposed here, normally distributed perception errors for the unguided drivers are used, but with a standard deviation of θc_a^{UE}, where c_a^{UE} is the travel cost for link a corresponding to a (deterministic) user equilibrium flow pattern for all drivers. This is preferred because it is more closely related to the idea that larger perception errors are made with larger travel times and greater congestion, rather than using the free-flow travel time which may be more related to the physical characteristics of the link (for example, an uncongested freeway would have a relatively large free-flow travel time and would thus counter-intuitively tend to result in large perception errors). The guided SUE drivers are modeled in the same way, except that their link travel time standard deviation is Ψc_a^{UE}, where $0 < \Psi < \theta$ (i.e., guidance tends to reduce the size of the errors made by equipped drivers). The errors are distributed independently among user classes and among links. It is noted that this model is somewhat unrealistic in one respect, because the journey time prediction methods will tend to be more precise with larger levels of take-up—data on actual travel times relayed to the guidance system from the beacons will relate only to equipped vehicles, and so an increase in level of take-up will essentially lead to an increase in sample size. It would be expected, then, that the variance of the random errors would be a decreasing function of the level of take-up. Because no suitable relationship of this kind was available, however, it was necessary to retain the assumption of a constant error variance relative to the proportion of vehicles equipped.

Finally, Daganzo proposes a solution algorithm for the multiple user class equilibrium problem with stochastic costs, which is essentially an extension of the method of successive averages, as introduced by Sheffi and Powell (*15*). Convergence of this algorithm to the equilibrium solution is guaranteed for the cost functions considered here. Daganzo's scheme was implemented as follows:

1. Set $f_{ai}^{(0)} = 0$, a, i, where $f_{ai}^{(r)}$ refers to the estimate of the equilibrium user class flows at iteration r. Set $r = 0$.
2. Calculate $\underline{F}^{(r)}$ from

$$\underline{F}^{(r)} = \sum_j f_{aj}^{(r)}$$

and hence the costs $c_{ai}^{(r)}$ corresponding to $\underline{F}^{(r)}$ ($\forall\ a, i$).

3. For each user class i
   (a) Sample a set of link error terms θ_{ai} ($\forall\ a$) from the specified probability distribution, by a pseudo-randomization process, and set

$$C_{ai}^{(r)} = c_{ai}^{(r)} + \theta_{ai} \forall\ a$$

   (b) Perform an all-or-nothing assignment for this user class using the randomized costs $C_{ai}^{(r)}$—yielding a set of user class link flows $g_{ai}^{(r)}$ ($\forall\ a$)
   (c) Set

$$f_{ai}^{(r+1)} = (1 - 1/r)f_{ai}^{(r)} + 1/rg_{ai}^{(r)} \ (\forall\ a)$$

   (d) Set

$$F_a^{(r)} = F_a^{(r)} - f_{ai}^{(r)} + f_{ai}^{(r+1)} \ (\forall\ a)$$

and recompute the costs corresponding to $\underline{F}^{(r)}$—store again in $c_{ai}^{(r)}$ ($\forall\ a$)

(4) Set $r = r + 1$ and return to Step 2 until the predetermined number of iterations are complete.

The main advantage of the guidance strategy described in the preceding may be seen to be that it anticipates the effect of rerouted traffic on travel times in the network, and in this way, directs traffic to one or more routes. It would be hoped that such an approach would be effective even with a high proportion of equipped vehicles (and this will be investigated for the test networks studied here). The single route strategies currently being considered for field trials [Von Tomkewitsch (*18*) and Belcher and Catling (*19*)] are likely, on the other hand, only to be of use when the proportion of equipped vehicles is small, so that rerouted traffic has only a relatively small effect on delays [see the earlier comments on the findings of Breheret et al. (*7*)]. Increasing the frequency at which such single route strategies are updated as the level of equipped vehicles increases is clearly one possibility, but there are limitations on this frequency imposed, for example, by the time it takes to communicate information from the beacons to the guidance system. In the future, then, as the popularity of route guidance grows, multiple route strategies are likely to be an essential component of the guidance system.

A second notable feature of the model is that it supposes unguided drivers will respond to the new routes taken by the equipped vehicles, and will aim to choose new routes which minimize their personal (perceived) travel cost. Given that an equilibrium-based assignment is accepted here as a reasonable approximation to the long-term average route choice under steady state conditions, then the assumption seems natural that unguided drivers will—in the long run—still seek minimum cost routes when guided drivers are in the network. It is recognized, however, that this is a strong behavioral assumption, neglecting any loyalty drivers might have had to their chosen routes before guided drivers were introduced [compare with Mahmassani and Jayakrishnan (*14*)], and would in any case be much more difficult to justify in situations where a frequently updated dynamic route guidance system was in operation.

TEST RESULTS

The guidance strategy was implemented using an adaptation of the simulation model SATURN (*20*) and the solution algorithm of Daganzo, as described previously. The two real-life networks considered were those of Weetwood (an area of Leeds) consisting of 70 zones, 104 intersections, and 440 links; and of Barcelona consisting of 110 zones, 820 intersections, and 2,547 links. The cost functions used were of the form of Equation 1, where k was given the value 5 for both networks. (The two networks had been calibrated originally using different powers in the cost functions on different links—the value 5 was chosen as it was approximately the average of all these powers, in both cases).

For each network, the guidance model was implemented under the following:

1. Three different demand levels: 100 percent, 130 percent, and 160 percent of the observed origin-destination flows, cor-

responding to an average network speed (before guidance) of approximately 15, 25, and 35 km/h, respectively;

2. Nine different levels of equipped vehicles: 0 percent, 5 percent, 10 percent, 20 percent, 30 percent, 50 percent, 70 percent, 90 percent, and 100 percent; and

3. Three different routing criteria: with equipped drivers either all guided as a UE, all guided as a SO, or all guided as a SUE (with two different levels of error in this latter case).

Finally, to decide upon a suitable value for a parameter θ, which determines the link travel time variances for the unguided drivers, an idea from Breheret et al. (7) is used. For a number of values of θ, the average inefficiency $I(\theta)$ is calculated by

$$I(\theta) = 100\,(p(\theta) - 1)\ \%$$

where

$$p(\theta) = \frac{\text{Total system travel time under SUE}(\theta)}{\text{Total system travel time under UE}}$$

and where SUE(θ) means an SUE assignment for the whole origin-destination matrix, with parameter θ. That is, assuming that drivers are aiming to follow a UE, $I(\theta)$ is a measure of the average excess travel time incurred by their perception errors.

For the purposes of this paper, for a given network, a value for θ is then chosen which gives rise to an inefficiency of the order of 5 to 6 percent for each of the demand levels considered. The reasoning behind this is that various studies have shown that the percentage wastage caused by drivers not fulfilling their objective of choosing the minimum time or minimum distance route is of this order—for example, Lunn (21) estimated the average excess on all journeys in Great Britain, excluding commuter trips, to be at least 5 percent of total costs; Wootton et al. (22) arrived at figures of 4 to 6.5 percent inefficiency; and Jeffrey (13), from analyzing times and distances corresponding to a sample of journeys made in the United Kingdom, concluded that the average inefficiency of drivers was around 6 percent. The use of inefficiency to build a suitable route choice model for unguided drivers is appealing, in that it allows the calibration of the model against observed data (though in a very coarse manner).

It should be emphasized here that although the use of inefficiency in this way appears to be a sensible one, it has the disadvantage that in specifying a value for θ, the system benefit of UE guidance at 100 percent take-up is directly determined, and the benefits of other scenarios are clearly greatly affected by the value chosen. It means therefore that the model should not be used to infer absolute measures of the effects of route guidance; its purpose is to compare the effects of different levels of take-up, demand levels, routing strategies, and so on.

The values of I(θ) for a number of values of θ are given in Figures 1 and 2. For Weetwood, there is a clear pattern of an increased I(θ) with increased θ, or greater demand. The value $\theta = 0.3$ is chosen for the purposes of further investigation, giving an average inefficiency of 6 percent over the three demand levels. For the Barcelona network, the pattern is somewhat different, with much less difference between demand levels and with the possibility of I(θ) decreasing with greater demand (demand level 1 showing greatest inefficiencies). It is still the case, however, that I(θ) is an increasing function of θ. The value of $\theta = 0.4$ is chosen for future study.

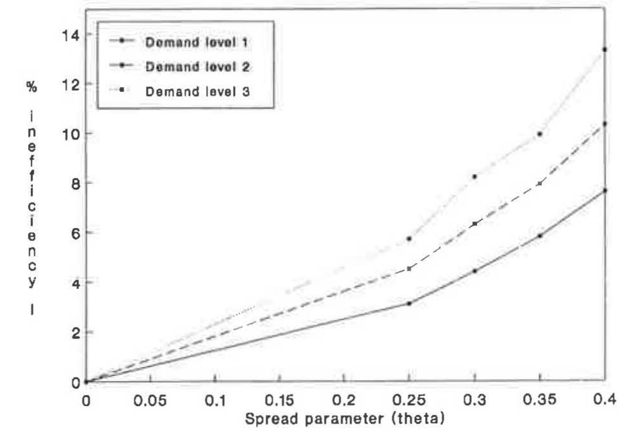


FIGURE 1 Inefficiency for Weetwood.

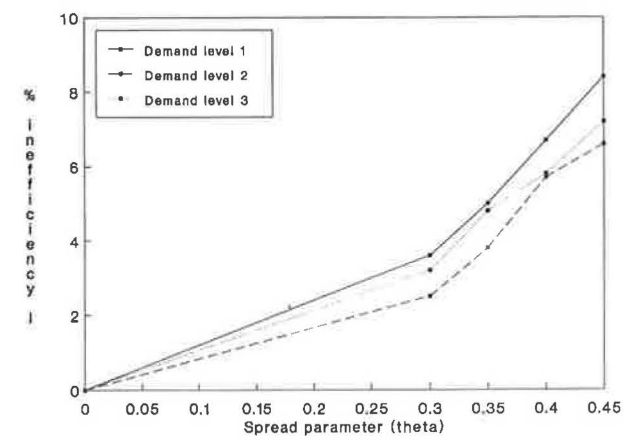


FIGURE 2 Inefficiency for Barcelona.

There are two studies with which some comparison may be drawn on this point of modeling unguided drivers. Breheret et al. (7), in using a uniform error structure for perceived costs (with flow dependent range), found the relationship between inefficiency and spread parameter to be highly network dependent and demand dependent—because of this, and because they give no indication as to the size of the networks or the absolute levels of congestion, it is difficult to draw any further parallels with this work. Koutsopoulos and Lotan (11), on the other hand, used a very similar "before guidance" model to that considered here, the most notable disparity being their use of a flow dependent perception error variance of $\phi c_a(F_a)$. For their study on a network of a similar size to Weetwood, they used a value of $\phi = 0.5$, which gave rise to an inefficiency of around 4 percent (relative to the UE, $\phi = 0$, case) for the three demand levels considered.

It is evident that quite large values of the spread parameters are required to give realistic inefficiencies. In one respect this is unappealing, since—as it makes sense to truncate the perceived travel time distributions at zero—the randomization may be biased (although for the values considered here the bias will tend to be very small). A factor of greater concern is that in using the approach proposed in this paper to model unguided drivers, some of the deficiencies of the model in reproducing a realistic route choice criterion are contained in θ. There is part of the inefficiency, therefore, which cannot be recovered by route guidance. In routing all guided drivers as a SUE, a third component (in addition to errors in the information used by the guidance system and the nonadherence of equipped drivers to the recommended routes), which contributes to the error terms, may be regarded as lack-of-fit of the model to the observed value of 6 percent inefficiency—the observed value was calculated by comparing minimum time routes with what drivers actually did, and so also accounts for drivers who were not intending to follow minimum time routes.

The model was applied to the networks described, with unguided drivers modeled by a SUE with $\theta = 0.3$ for Weetwood and $\theta = 0.4$ for Barcelona. From initial studies of the Weetwood network, it appeared that to obtain a reasonable degree of convergence, a large number of iterations would be required—the stopping criterion chosen was the completion of 200 iterations. Although this may have been the case too for Barcelona, the size of the network meant that such a large number of iterations would be computationally prohibitive, and so only 30 iterations were carried out in this case. The results are given in Figures 3 to 14. In the figures on

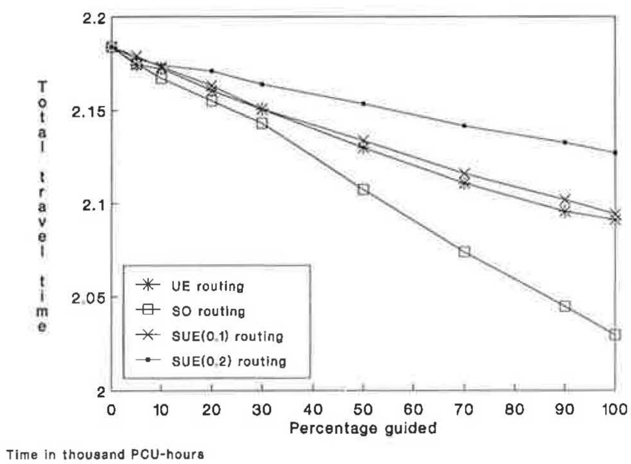


FIGURE 3 System benefit for Weetwood demand level 1.

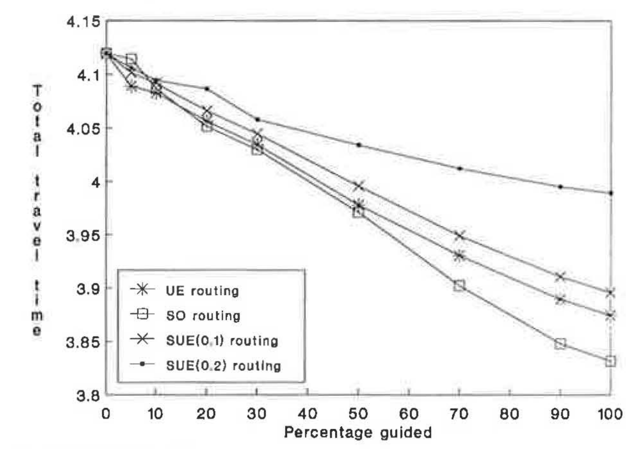


FIGURE 4 System benefit for Weetwood demand level 2.

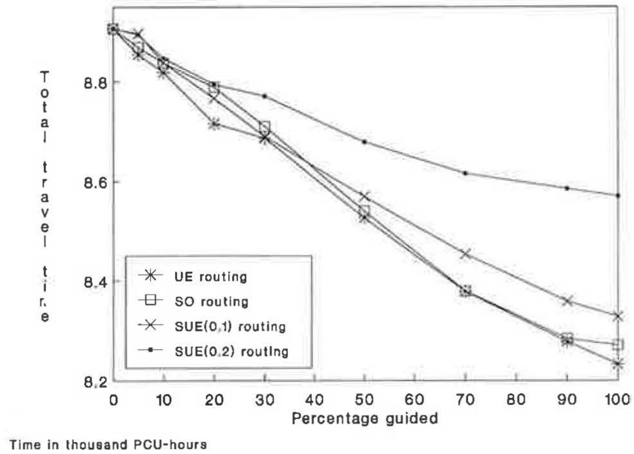


FIGURE 5 System benefit for Weetwood demand level 3.

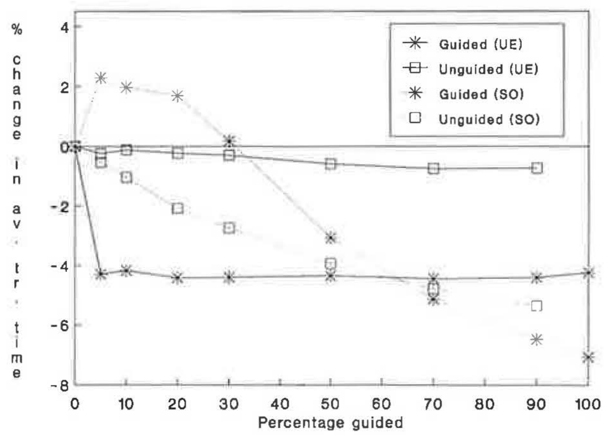


FIGURE 6 Individual benefits for Weetwood demand level 1.

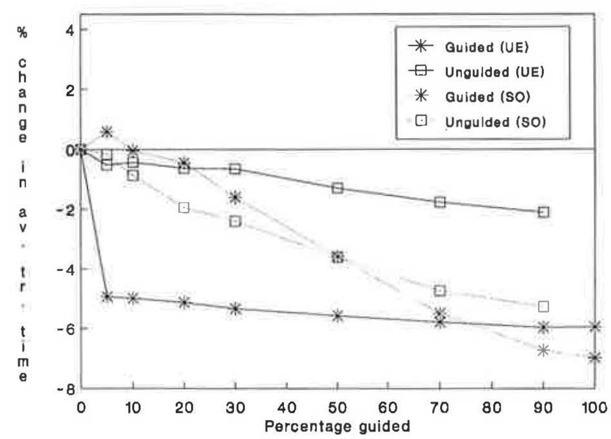


FIGURE 7 Individual benefits for Weetwood demand level 2.

system benefit (expressed in terms of total travel time, Figures 3 to 5 and 9 to 11), each of the four strategies is represented—for example, the trend labeled SUE(0.2) refers to the strategy in which all equipped vehicles are guided according to a SUE with $\Psi = 0.2$. In the figures on individual benefits (Figures 6 to 8 and 12 to 14), results relating to the UE (broken line)

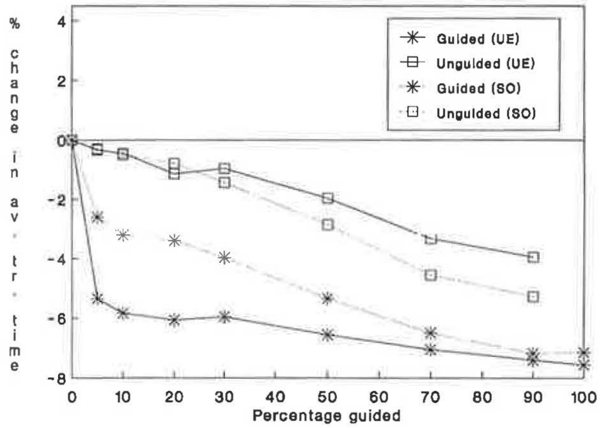
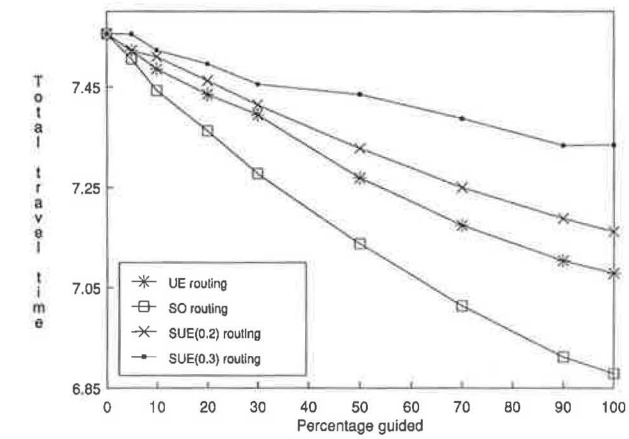
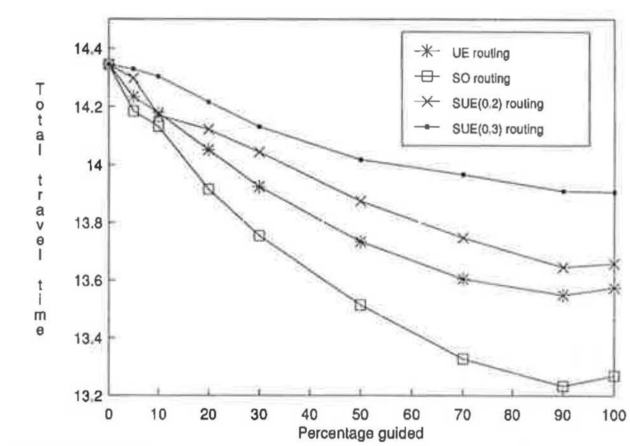


FIGURE 8 Individual benefits for Weetwood demand level 3.



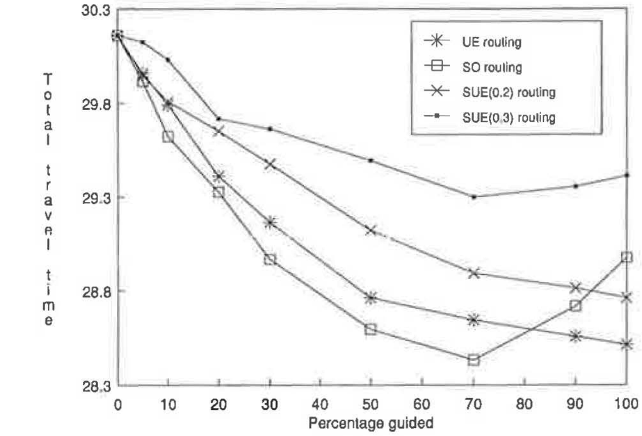
Time in thousand PCU-hours

FIGURE 9 System benefit for Barcelona demand level 1.



Time in thousand PCU-hours

FIGURE 10 System benefit for Barcelona demand level 2.



Time in thousand PCU hours

FIGURE 11 System benefit for Barcelona demand level 3.

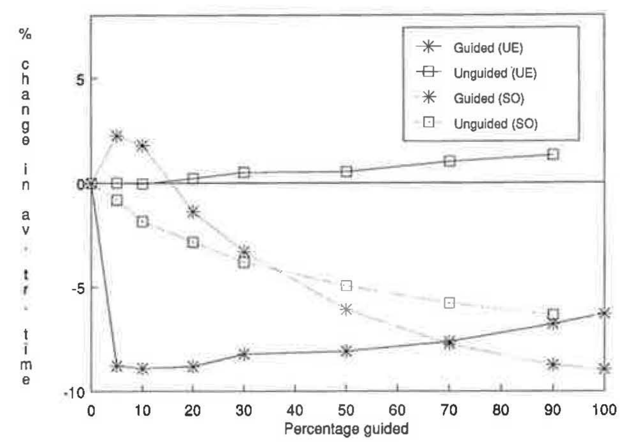


FIGURE 12 Individual benefits for Barcelona demand level 1.

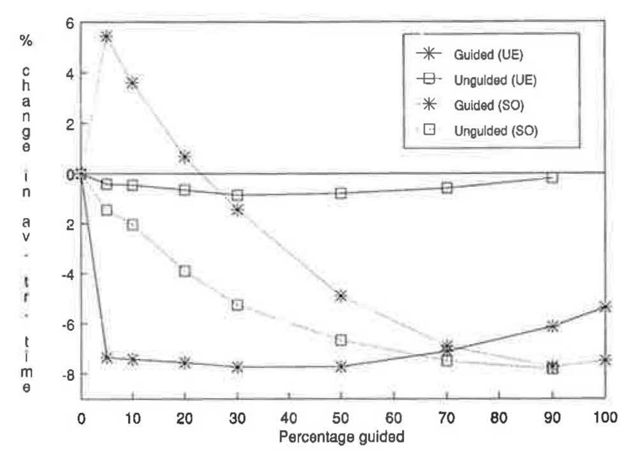


FIGURE 13 Individual benefits for Barcelona demand level 2.

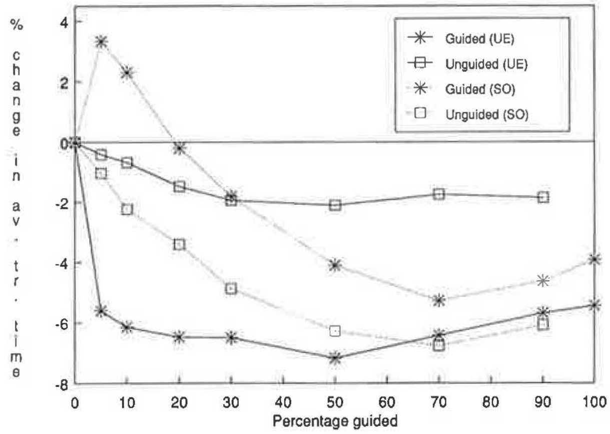


FIGURE 14 Individual benefits for Barcelona demand level 3.

and SO (continuous line) routing strategies only are given, with each separated into guided (star symbol) and unguided (square symbol) vehicles. So that, for example, in Figure 6 the label "Guided (UE)" refers to guided drivers under the strategy in which all equipped drivers are guided as a UE—that is, case A in Table 1—and "Unguided (UE)" refers to unguided drivers under the same strategy.

The first point to note is that in a small number of situations, there is evidence of strange behaviour—first, in Figure 5, for Weetwood at demand level 3 with 100 percent of vehicles equipped, the total travel time arising from SO routing is greater than that arising from UE routing, which clearly should not be the case. For Barcelona at higher demand levels (Figures 10 and 11) a similar problem is evident—for example, in Figure 11, an assignment with 70 percent guided according to a SO routing gives a smaller total travel time than a 100 percent SO routing. This point will be addressed further in the next section.

For the system benefit (as measured by total travel time), it can be seen that for all of the routing strategies considered and for both networks, guidance offers an improvement over the base (no guidance) situation, for all levels of equipped vehicles and all demand levels considered. In all cases, the travel time saving becomes greater as the level of take-up increases (with one or two exceptions—see preceding comments on convergence), following an almost linear trend in the Weetwood case. Below a 50 percent take-up, the percentage saving in total travel time tends to be higher with higher demand levels, although in all cases the differences between demand levels are not great. Concentrating specifically on the UE and SO routing strategies, Figures 3 to 5 show that for Weetwood, the percentage saving in total travel time due to guidance increases with demand for most levels of take-up—for example, between 10 and 90 percent take-up, UE routing gives savings of 0.5–4.0 percent, 0.9–5.6 percent, and 1.0–7.1 percent, respectively, for demand levels 1, 2, and 3; SO routing, on the other hand, gives respective savings of 0.7–6.4 percent, 0.8–6.6 percent, and 0.8–7.0 percent. The differences between demand levels are, however, not great, and the same is true for the Barcelona network. In this latter case, though, there is a slight decrease in the benefits attainable at higher levels of take-up as demand increases (Figures 9 to 11)—the savings for 10–90 percent take-up are 0.9–6.0 percent, 1.2–5.5 percent, and 1.2–5.3 percent under UE routing for demand levels 1, 2 and 3 respectively, and under SO routing the respective savings are 1.5–8.5 percent, 1.5–7.7 percent, and 1.8–5.8 percent. On the whole, for both networks, the pattern is as one may expect, with an increase in total travel time savings as decreases for SUE routing (down to $\Psi = 0$ for UE routing), all of these giving rise to larger total travel times than SO routing.

For the benefit to individuals (the percentage decrease in average travel time with the guidance system in operation), it may be seen that for a UE routing, equipped drivers are always better off with guidance. A striking feature is that this benefit is approximately constant for all levels of take-up—notably, the benefit is achieved at a very low percentage of equipped vehicles (in fact at 1 percent take-up, not shown in the graphs), and does not decrease notably at higher participation levels. From Figures 6 to 8, it may be seen that under UE routing, guided drivers save around 4 percent, 5 percent, and 6 percent respectively for Weetwood demand levels 1, 2 and 3, whereas (Figures 12 to 14) the savings are around 8 percent, 7 percent, and 6 percent for Barcelona demand levels 1, 2 and 3. Under such a routing scenario, the change in travel time for unequipped drivers is always small relative to the benefit to guided drivers (always less than 3 percent, and usually less than 1 percent, with an actual disbenefit for the Barcelona network at demand level 1), although there appears to be slightly greater benefit to them as congestion increases from demand level 1 to demand level 3. In most of the situations, the benefit to individual unguided drivers also tends to increase with level of take-up.

With SO routing, there is a disbenefit to individual guided drivers on average, for lower levels of take-up (low being levels of equipped vehicles less than of the order of 10 to 30 percent); for higher levels of take-up, on the other hand, guided drivers experience a saving in travel time which increases with level of take-up. Above 50 percent take-up the savings for equipped drivers under SO routing are 3 to 7 percent for Weetwood and 5 to 9 percent for Barcelona, depending on level of take-up and demand level. The journey time saving for unguided drivers tends to be somewhat larger here than with UE routing, particularly at lower levels of take-up, with guided and unguided drivers benefiting similar amounts at higher levels of take-up.

Comparing the UE and SO routing strategies for both networks, it may be seen that UE guidance will always benefit the equipped drivers most, with only limited benefits to the unequipped drivers, but giving rise to considerable system benefits. Not surprisingly, SO routing primarily benefits the unguided drivers—at the expense of the guided drivers—at lower levels of equipped vehicles. However, equipped drivers start benefiting too when their numbers increase. At the highest levels of equipped vehicles (more than 50 to 70 percent), under such a routing strategy the guided drivers may even benefit more than the unguided drivers, despite being guided to minimum marginal cost routes. The system benefits of SO routing are higher than with UE routing, but may not warrant the disbenefits to equipped drivers at lower participation levels.

In this paper, there has only been space to discuss a relatively small number of features of the guidance model; in various other papers (*23–25*), a description of many other aspects is given, such as the effect on distance traveled; the performance in situations of unforeseen variability; the influence of the reactions of unguided drivers; properties of the recommended routes; and guided drivers' acceptance of SO advice.

CONVERGENCE

The simulations reported in this paper have used the completion of a specified number of iterations of the solution algorithm as a stopping criterion. In this section, the aim is to briefly investigate the influence of this stopping criterion, particularly with respect to the problem results identified earlier. As examples, Barcelona demand level 3 (100 percent take-up) and Weetwood demand level 2 (5 percent take-up) will be considered under UE and SO routing.

Although convergence indicators do not appear to have been proposed for use with Daganzo's algorithm, some work has been done in this area with the standard method of successive averages in the one user class SUE case. It is noted that a choice of indicator on the basis of similarity of link flows between iterations is not altogether straightforward, because the search direction is a random variable and the step size ($1/r$) is fixed (unlike, for example, the Frank-Wolfe algorithm for the deterministic UE problem)—a good discussion of this point is given by Sheffi (*17*). Sheffi and Powell (*15*) have, however, proposed an indicator I_r for this problem:

$$I_r = \sum_a s_a^{(r)} / \sum_a F_a^{(r)}$$

where for each link a, $F_a^{(r)}$ and $s_a^{(r)}$ are, respectively, the sample mean and sample standard deviation of the flow estimates for that link on the last m iterations, for some user-specified m. The indicator was used as a stopping criterion ($I_r < 0.0005$, with $m = 5$) for Barcelona demand level 3 with 100 percent take-up, for each of UE and SO routing; the results at convergence were:

	No. of iterations	*Total travel time*
UE	25	28,836.0
SO	33	28,692.9

The SO strategy now gives rise to a smaller total travel time than the UE one, needing a larger number of iterations to reach the same degree of convergence. Extending the definition of I_n to the multiple user class case was then achieved by applying it to the *total* link flows; this makes more sense than studying the user class flows separately, since our results only guarantee uniqueness of the total link flows. For the Weetwood network with 5 percent guided ($m = 5$, $I_n < 0.00005$), the results obtained were:

	No. of iterations	*Total travel time*
UE	217	4,088.5
SO	203	4,113.4

and the progress of the algorithm is illustrated in Figure 15 (with respect to the indicator I_n and Figure 16 (with respect to total travel time). Although total travel times again appear to suggest a slower convergence for the SO strategy than the UE strategy, this is not apparent in the indicator I_n, and in fact the algorithm under SO routing is terminated *before* that under UE routing. The same problem encountered with the fixed number of iterations case is evident here, with SO routing giving rise to a higher total travel time than UE guidance. Because of this inconsistency in the performance of the indicator (other scenarios tested reinforced this inconsistency), it was decided to retain the stopping criterion of a fixed number of iterations, although this is clearly an area in which more work is needed.

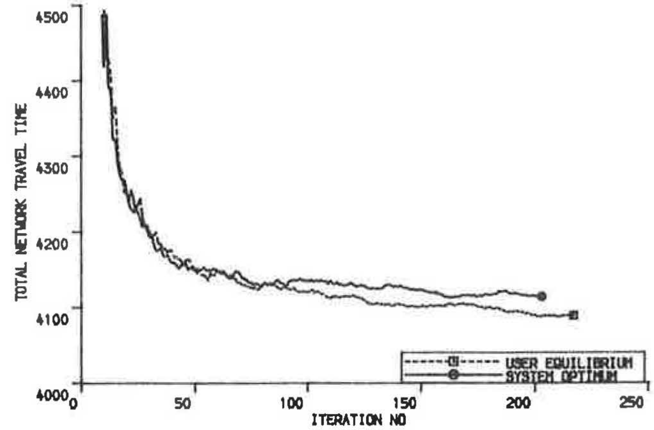


FIGURE 15 Convergence with respect to indicator, I_r (Weetwood, 5 percent guided, demand level 2).

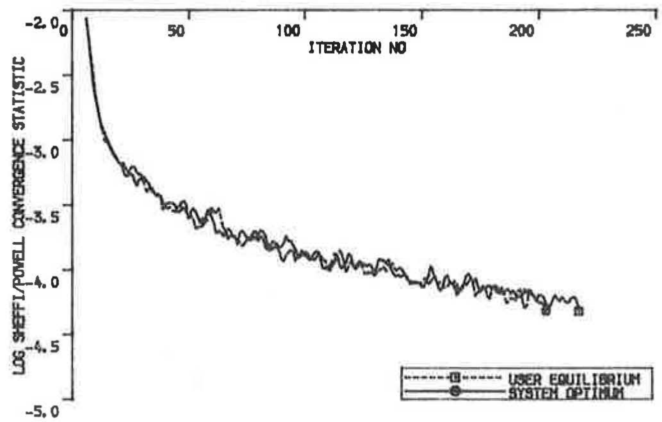


FIGURE 16 Convergence with respect to total travel time (Weetwood, 5 percent guided, demand level 2).

CONCLUSION

A model of a route guidance system has been proposed in terms of a multiple user class equilibrium assignment, with vehicles divided into equipped and unequipped classes, the former being subdivided further dependent on the routing criterion used and the quality of the information supplied. Guidance is used to route vehicles either to a user equilibrium or to a system optimum flow pattern, assuming that without guidance drivers aim to follow a user equilibrium but fail to do so because of perception errors in their evaluation of travel times. Furthermore, unequipped drivers are assumed to respond to the new route choice of guided drivers, and seek a new user equilibrium routing.

For cost functions of a particular polynomial form, it is shown that such routing strategies, in combination with the route choice of unequipped drivers, are guaranteed to lead to a unique and stable equilibrium flow pattern with respect to total link flows.

The main advantage of such an equilibrium-based strategy is that it spreads the traffic between multiple routes on each origin-destination movement, and so would be expected to lead to effective guidance even when a high proportion of drivers are equipped with a guidance device. The test runs on two real-life networks were used to investigate such a property, as well as the performance of the strategies in several different scenarios. It was seen that both user equilibrium and system optimal guidance generally reduced the total system travel time, the benefit being an increasing function of the level of take-up. The level of congestion appeared to have less effect on the benefits than the percentage guided, although there was some indication of slightly greater percentage savings in more congested situations below 50 percent take-up.

Under UE routing, individual guided drivers experienced a significant reduction in average travel time, this being approximately constant (on the order of 5 percent, varying with demand level) for all levels of take-up. Unguided drivers also tended to benefit from such guidance, but always to a much lesser degree than guided drivers (usually less than 1 percent reduction in average travel time); their savings tended to increase as the percentage of equipped drivers or the level of congestion increased.

SO routing was found to lead to a slightly greater reduction in total travel time than UE routing, particularly in the least congested scenarios. The effects on individual groups of drivers are, however, quite different in the two cases. SO routing was seen to primarily benefit unequipped drivers, significantly improving their position in comparison with UE routing. At lower levels of take-up (10 percent or less) this saving tends to be at the expense of equipped drivers, who may experience an increase (of as much as 5 percent in the extreme) in average travel time due to guidance. For higher levels of take-up, equipped drivers will benefit too from SO routing, with the saving in average travel time growing in similarity (lending to the order of 5 percent) as the level of take-up increases.

In terms of future research, if the model proposed here is to be used further there is a need to determine a suitable convergence indicator for the solution algorithm—possibly on the basis of development of an objective function for the problem, or on a "gap" function monitoring route-based properties—to overcome the problem of apparently different convergence rates of UE and SO routing.

There is also a need for a better modeling both of unguided drivers—in particular, their response to the implementation of route guidance—and of guided drivers, with quite different reactions likely to occur when system optimal as opposed to user equilibrium routing is recommended. Furthermore, relationships are needed between the level of take-up of guidance and the level of errors in the journey time prediction algorithms, the latter tending to decrease as the sample from which the prediction is made (i.e., the guided drivers) increases in size.

Taking a wider view, the limitations of the model used here have to be recognized; the strategies have been assessed in a steady-state framework, under long-term average network conditions and demand. Under a rolling program of research funded by the Science and Engineering Research Council of Great Britain, *Fundamental Aspects of Full-Scale Dynamic Route Guidance*, one strand of the work being carried out at Leeds is a fundamental assessment of network models, with an emphasis on the representation of dynamics, variability, and driver behavior. The specification, and subsequent design, of such a model is seen as an important stage in the future evaluation of route guidance systems.

ACKNOWLEDGMENTS

This work was funded by the EC DRIVE research program and by the Science and Engineering Research Council of Great Britain.

REFERENCES

1. S. C. Dafermos. The Traffic Assignment Problem for Multiclass User Transportation Networks. *Transportation Science*, Vol. 6, 1972, pp. 73–87.
2. D. Van Vliet, T. Bergman, and W. H. Scheltes. Equilibrium Traffic Assignment With Multiple User Classes. Proceeding of PTRC Summer Annual Meeting, Seminar M, 1986, pp. 111–122.
3. F. Kobayashi. Feasibility study of route guidance system. *Transportation Research Record 737*, TRB, National Research Council, Washington, D.C., 1979, pp. 107–112.
4. H. Tsuji, R. Takahashi, H. Kawashima, and Y. Yamamoto. A Stochastic Approach for Estimating the Effectiveness of a Route Guidance System and Its Related Parameters. *Transportation Science*, Vol. 19, 1985, pp. 333–351.
5. H. Al-Deek, M. Martello, A. D. May, and W. Sanders. Potential Benefits of In-Vehicle Information Systems in a Real-Life Freeway Corridor Under Recurring and Incident-Induced Congestion. Proceedings of 1st Vehicle Navigation and Information Systems Conference, Toronto, Canada, 1989, pp. 288–291.
6. R. Rakha, M. Van Aerde, E. R. Case, and A. Ugge. Evaluating the Benefits and Interactions of Route Guidance and Traffic Control Strategies Using Simulation. Proceedings of 1st Vehicle Navigation and Information Systems Conference, Toronto, Canada, 1989, pp. 296–303.
7. L. Breheret, N. B. Hounsell, and M. McDonald. The Simulation of Route Guidance and Traffic Incidents. Presented at 22nd Annual Universities Transport Studies Group Conference, Hatfield, UK, (1990).
8. D. R. Leonard, J. B. Tough, and P. C. Baguley. *CONTRAM—a Traffic Assignment Model for Predicting Flows and Queues During Peak Periods*, TRRL Report LR 841, Transport and Road Research Laboratory, Crowthrone, UK, 1978.
9. J. C. Smith and K. Russam. Some Possible Effects of Autoguide on Traffic in London. Proceedings of 1st Vehicle Navigation and Information Systems Conference, Toronto, Canada, 1989, pp. 282–287.
10. T. Van Vuren, D. Van Vliet, and M. J. Smith. Combined Equilibrium in a Network With Partial Route Guidance. In *Traffic Control Methods* (S. Yagar and S. E. Rowe, eds.), Engineering Foundation, New York, 1989, pp. 375–387.
11. H. N. Koutsopoulos and T. Lotan. Effectiveness of Motorist Information Systems in Reducing Traffic Congestion. Proceedings of 1st Vehicle Navigation and Information Systems Conference, Toronto, Canada, 1989, pp. 275–281.
12. C. F. Daganzo. Stochastic Network Equilibrium With Multiple Vehicle Types and Asymmetric, Indefinite Link Cost Jacobians. *Transportation Science*, Vol 17, 1983, pp. 282–300.

13. D. Jeffery. Route Guidance and In-Vehicle Information Systems. In *Information Technology Applications in Transport*, Utrecht, V.N.U. Science Press, 1987, pp. 319–351.
14. H. S. Mahmassani and R. Jayakrishnan. System Performance and User Response Under Real-time Information in a Congested Traffic Corridor. Presented at Seminar on Urban Traffic Networks, Capri, Italy, 1989.
15. Y. Sheffi and W.B. Powell. An Algorithm for the Equilibrium Assignment Problem With Random Link Times. *Networks*, Vol. 12, 1982, pp. 191–207.
16. C. F. Daganzo. Unconstrained Extremal Formulation of Some Transportation Equilibrium Problems. *Transportation Science*, Vol. 16, 1982, pp. 332–360.
17. Y. Sheffi. *Urban Transportation Networks*. Prentice-Hall, Englewood Cliffs, N.J., 1985.
18. R. Von Tomkewitsch. ALI-SCOUT: A Universal Guidance and Information System for Road Traffic. Presented at 6th IEE International Conference on Automotive Electronics, London, England, 1987.
19. P. L. Belcher and I. Catling. Electronic Route Guidance by AUTOGUIDE: the London Demonstration. *Traffic Engineering and Control*, Vol. 28, 1987, pp. 586–592.
20. D. Van Vliet. SATURN—a Modern Assignment Model. *Traffic Engineering and Control*, Vol. 23, 1982, pp. 578–581.
21. S. E. Lunn. *Route Choice by Drivers*. Department of Environment; Department of Transport, Transport and Road Research Laboratory, Supplementary Report 374, 1978.
22. H. J. Wootton, M. P. Ness, and R. S. Burton. Improved Direction Signs and the Benefits for Road Users. *Traffic Engineering and Control*, Vol. 22, 1981, pp. 264–268.
23. D. P. Watling. *Route Guidance Algorithms Effective for all Levels of Take-up and Congestion*. Working paper 315, Institute for Transport Studies, University of Leeds, 1990.
24. D. P. Watling. *A Comparison of System Optimal and User Optimal Route Guidance*. Working Paper 316, Institute for Transport Studies, University of Leeds, 1990.
25. D. P. Watling. *System Optimal Versus User Optimal Route Guidance*. Presented at Universities Transport Study Group Annual Conference, University of Nottingham, UK, Jan. 1991.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

New Dynamic Model to Evaluate the Performance of Urban Traffic Control Systems and Route Guidance Strategies

M. J. SMITH AND M. O. GHALI

A dynamic assignment/control model for evaluating urban traffic control schemes and route guidance strategies is described in this paper. The model, which has been developed recently at the University of York, United Kingdom, uses the dynamic assignment program CONTRAM. For each fixed road network, subject to a given overall pattern of dynamic demand, the model currently evaluates all eight possible combinations of (*a*) four responsive traffic control policies, and (*b*) two route-guidance strategies. The model assumes that the route-guidance is accurately obeyed. The model gives results for all required demand levels. Preliminary results of applying the model to one artificial network and one realistic network are also presented. Further extensions are in hand to take some account of (*a*) drivers' variable compliance with guidance, (*b*) the effects of different traffic signals following different control policies, and (*c*) drivers' ability to select their own departure time. At the moment, the model requires that the road network is fixed, and does not seek to mimic the dynamics associated with the beginning and end of an incident. Good communications with drivers will help to ensure that the guidance advice is obeyed; but for system-optimal-like guidance strategies a supportive, well-communicated, road pricing system may also be necessary to obtain such obedience. The model described here will clearly be useful for assessing the consequences of different road pricing strategies in a dynamic setting.

A dynamic assignment model to evaluate traffic control systems and route guidance strategies is needed. The performance of urban road traffic networks is of major importance for travelers, nontravelers, transportation engineers, and town planners. Yet there is currently no computer model which is trusted to provide realistic estimates of traffic network performance.

Current simulation programs (such as NETSIM) accurately represent small-scale interactions but do not model route-choice decisions and, on the other hand, most current assignment programs do not have accurate, dynamic, simulations of the detailed interactions between traffic flows and signal controls at junctions; nor do they allow a range of control systems.

It is clear that, in reality, network performance depends on both network-wide properties and on the intimate detail of particular junctions including the effects of traffic control, at least in part because both affect the routing decisions of drivers. So there is no escape from the need to somehow estimate the networkwide effect of local interactions (including responsive control and drivers' route choices) if network performance is to be accurately estimated.

Department of Mathematics, University of York, Heslington, York YO1 5DD, United Kingdom.

Recent attempts to study the benefits of route guidance have made this need more clear and more urgent; the European DRIVE project is now sharply feeling the lack of a reliable, reasonably fast, network model which can test various control and routing strategies in a realistic, dynamic, setting.

While the need for a good dynamic assignment model has been highlighted by the particular current wish to evaluate dynamic route guidance strategies, such a model is also essential if we are to evaluate responsive traffic control systems even if route guidance is not involved. Responsive control systems, without route guidance, are becoming increasingly sophisticated and increasingly used; a thorough scientific evaluation of these systems is very long overdue. The only way to perform such an evaluation is by using a good, flexible, dynamic, control/assignment program.

Mahmassani et al. (*1*) acknowledge the difficulty of assessing the likely performance of a real urban traffic network; they suggest that "the complex interactions in a traffic network effectively preclude the analytic derivation of macroscopic network-level relationships from the basic principles governing microscopic traffic behaviour . . . ." Mahmassani et al. proceed to outline an approach to network modeling using Herman and Prigogine's (*2*) two-fluid theory of town traffic, and give some results of simulation experiments. These authors acknowledge the limited nature of their experiments, which do not allow for either drivers' route choices or the detailed influences possible with modern responsive control systems.

A new dynamic model for assessing traffic control system and route guidance strategies is described in this paper. This model has been obtained by extending the dynamic assignment program CONTRAM (*3–5*), developed in 1978 by the Transport and Road Research Laboratory in the United Kingdom.

THEORETICAL BACKDROP

The need to take account of the effect of control on assignment was pointed out by Allsop (*6*). Smith (*7–10*), Smith et al. (*11*) and Smith and Ghali (*12*) have studied the theory involved when control and assignment are combined, and Heydecker (*13*) has pointed to some difficulties which arise when detailed models of junctions are incorporated into an assignment model. Throughout this theory demand is supposed to be steady as time varies.

Smith and Ghali (*14*) present some control/assignment results for a gently rising rush hour and a new theoretical approach to dynamic assignment with fixed signals. The dynamic assignment theory is severely limited by the assumptions needed; nonetheless the authors are not as pessimistic as Mahmassani et al.

RECENT TECHNOLOGICAL ADVANCES AND THEIR IMPLICATIONS FOR NETWORK MODELS

Traffic control systems that respond cleverly and rapidly to existing traffic flows and queues are now capable of being provided. Using new communication methods, it is hoped that the ability to respond to the needs of individual travelers will come soon; by giving route guidance, or by choosing signal-settings which are appropriate for certain routes, for example. And the practicality of road pricing has increased dramatically as a consequence of "smart cards."

But these technological opportunities have not been matched either by a corresponding improvement in our understanding of road traffic networks or by corresponding developments in the computer modeling of traffic networks; there is a substantial risk that technological solutions to urban congestion will be designed with an inadequate theoretical background, and imposed on real-life networks without proper evaluation. Indeed it is likely that some technologically sophisticated traffic control schemes are now aggravating the congestion which they were designed to reduce! It would surely be best to reduce—or even eliminate—the possibility of this happening with route-guidance; by immediately (*a*) ensuring the development of a sound theory, and (*b*) developing a good dynamic traffic assignment program.

DESIGNING AND TESTING A TRAFFIC CONTROL SYSTEM

The obvious way to test the networkwide effect of an urban traffic control scheme is by using computer programs. The lack of an appropriate network model means that such tests are currently impossible; and so control systems are often first tested on the real-life road networks themselves. [See Holroyd, et al. (*15*), Holroyd and Robertson (*16*), Tarnoff (*17*), and Henry et al. (*18*), for example.] Such tests are expensive, and (at least when the network is congested) are difficult to interpret because of demand and other changes which, in real life, cannot be controlled, and have major effects. Moreover the results of these tests have often been negative, and sometimes *very* negative, at least when congestion is severe.

The need to control certain aspects of the problem even during computer tests has recently led to the creation of an option which fixes certain demand and routing aspects of NETSIM; without such options meaningful control experiments become difficult and essentially unrepeatable. Such options, fixing certain aspects of the problem while the effect of varying others is studied, cannot exist in real life!

Assuming that we wish to create a computer program which models a particular road traffic network with particular traffic signals and a specified dynamic demand; the central questions we seek to answer are:

1. For a given control response, what dynamic traffic flows will occur?
2. How should traffic signals respond to changing traffic?

Question 1 is a problem of prediction, given a specified control system operating in a specified way. Question 2 will also apply to a particular network equipped with particular control facilities; but this question is far more open-ended than Question 1; designing a good control response is far harder than predicting the outcome of a specified response. It is essential that Question 2 is answered if good control systems are to be designed.

DESIGNING AND TESTING A TRAFFIC CONTROL/ROUTE GUIDANCE SYSTEM

If dynamic route guidance is considered as a possibility for influencing road traffic, Questions 1 and 2 become

1′. For a given control response, and a given route-guidance strategy, what dynamic traffic flows will occur?

2′. How should traffic signals and routing advice respond to changing traffic?

If the signals and the route guidance are intended to react quickly to current traffic conditions, then Question 2′ requires a rapid response. This will be the case, in particular, if response to an incident is to be provided. None of these questions can be answered convincingly now because (*a*) the required theory is insufficiently developed, and (*b*) the tools necessary to study congested network performance, by using simulation on computers, do not exist.

The Sharp Questions

While the main function of Science is to explain and predict, an important function of mathematical models is to pose "sharp questions." Answers to Questions 1 and 1′ will clearly be fulfilling the main function to some degree—but Questions 2 and 2′ are sharp questions that require the careful development of a theory as well as the careful development of a rigorous dynamic model to test the theory. Sharp questions are important because they force one to try to get to the heart of the matter. The theory of traffic control and route choice in a dynamic setting is currently in a very underdeveloped state, and an appropriate dynamic network model involving both route-choice and traffic control does not exist.

Requirements of a Traffic Control/Route Guidance Network Model

For any specified network it would be sensible to use the network model so as to determine (traffic control policy, route-guidance strategy) pairs which produce good results, on the particular network, when taken together. [The new modification of the CONTRAM program allows this to some degree. As currently formulated, the program shows the consequences of a (control policy, route guidance) pair on the assumption, at the moment, that the route guidance is obeyed.]

If the model is really very fast, and includes a reasonable estimate of drivers' reaction to guidance, then the model could be used for on-line control and guidance; otherwise (and this is certainly the only hope at the moment) the model should be utilized off-line to generate a library of (traffic control policy, route guidance strategy) pairs and hence a library of (control, guidance) plans which are selected on-line in the light of current traffic conditions.

TWO EXISTING MODELS

There are two network models which have been used for assessing traffic management schemes in the United Kingdom. Both attempt to embrace both local detail and the networkwide assignment of traffic to routes according to reasonable route-choice principles; they are called CONTRAM and SATURN (*19*). Broadly speaking, SATURN captures the variety of local junction interactions better than CONTRAM, while CONTRAM has a more accurate representation of drivers' route-choice decisions in a dynamic, rush-hour, context. Of course, it is plain that in designing an assignment model which represents both junction detail and the dynamics of route-choice, some compromises have to be made. An extremely detailed dynamic assignment model with extremely precise local detail would, in the current state of the traffic modeling art, have an impossibly long running time on large realistic networks. In trying to improve upon current models it must be borne in mind that there are some theoretical problems which arise when accurate junction detail is included in a traffic assignment model: with certain junction delays assignment algorithms may not converge, and may also converge to different answers if the process is begun at different starting points; see Heydecker (*13*).

Neither CONTRAM nor SATURN, as they stand, is able to answer Question 2′ very well because both lack the flexibility to represent various control options conveniently and both have only one routing principle. Nonetheless both CONTRAM and SATURN have been used to test various route guidance scenarios; Breheret et al. (*20*) and Smith and Russam (*21*) used CONTRAM to assess, respectively, the likely benefit of (*a*) guiding drivers along their best routes accurately, and (*b*) using AUTOGUIDE in London. Van Vuren and D. Watling, in a paper in this Record, have used SATURN model to estimate the likely benefits of route guidance under a variety of circumstances. These studies both required detailed changes to the standard program.

In the work at the University of York CONTRAM was modified so that is can deal with a variety of control policies and a variety of route-guidance possibilities. It is hoped that the extended model will be used to test a whole spectrum of (traffic control policy, route guidance strategy) combinations. In this paper first results are given; the methods are more or less readily applicable to most real-life networks, although at this stage the model assumes that route guidance is obeyed.

Charlesworth (*22*) obtained similar control results in 1977 by linking a well-known steady-state assignment program to the TRANSYT signal optimization procedure. Although the dynamic assignment program CONTRAM and different signal optimization procedures are used in the present research, the basic idea is similar to that suggested by Allsop and employed in Charlesworth's paper.

CONTRAM

CONTRAM is a dynamic assignment model which represents, in a realistic manner, queueing delays as they vary in time and the routing changes which these time-varying delays are likely to cause. CONTRAM contains an option for setting the traffic signals according to a natural equisaturation policy, as well as a fixed-time option in which program users specify fixed signal settings. The program finds the dynamic flow pattern in which all drivers are using their least-time routes. This flow pattern will be called a user-equilibrium. Thus for any fixed signal-settings CONTRAM finds the user equilibrium consistent with these fixed settings; or alternatively CONTRAM finds the user-equilibrium consistent with a natural equisaturation policy. Having found the user equilibrium, CONTRAM prints out many figures which indicate the performance of the network. Much information about CONTRAM is contained in three reports, by Leonard et al. (*3*), Leonard and Gower (*4*), and Taylor (*5*) published by the Transport and Road Research Laboratory (in the United Kingdom).

New CONTRAM Extensions

In this paper we outline some extremely recent modifications and extensions of CONTRAM carried out at the University of York, United Kingdom. These allow the user to estimate the performance of traffic networks under a somewhat wider variety of (traffic control policy, route guidance strategy) combinations. We shall also give some results of applying our new version of CONTRAM to one artificial and one realistic network.

We have modified CONTRAM to accommodate

- A second route-choice strategy, and
- two further control policies.

The route-choice principle is that, perhaps because they are guided, drivers choose to travel along routes whose links have, in total, the least local marginal cost. Here the local marginal cost of traversing link i, beginning at some time t, is the additional cost caused to all users of this particular link by one additional vehicle entering this link, at this time; assuming that the signal-settings (and other drivers' route-choices) remain fixed. These routes do not cause least increase in the total travel time spent on the network; they are not routes of minimum marginal cost—they are routes of minimum local marginal cost. If the signal settings are fixed, the flow pattern which CONTRAM then calculates will be called a local system-optimal flow pattern for those fixed signal-settings.

The two further control policies are P_0 and local delay-minimization. The policy P_0 is introduced and discussed further in Smith (*7–10*). Operating in a steady-state context, the policy P_0 maximizes the capacity of a network (provided certain natural conditions are met).

The local delay-minimization policy minimizes delay at each junction for the dynamic flows which currently impinge on that junction. (The policy does not minimize delay for the network because it ignores two effects of a control change at the "current" junction. First, it ignores changes in the flows impinging on downstream junctions arising from changes in the throughput at the current junction. Second, it ignores routing changes as drivers react to changed travel times.) Our local delay-minimization policy chooses green-times which maximize throughput at each junction on the assumption that the junction is isolated.

Table 1 summarizes the (traffic control policy, route guidance strategy) combinations that can be studied with the program. Of these eight combinations only two come with the original CONTRAM. On the other hand, some of these combinations are unnatural and so would not be of much interest. We can also model mixtures in which some vehicles follow the user-equilibrium strategy and some follow the local system optimal strategy, and in which some junctions are controlled according to one control policy and some are controlled according to another. Furthermore, a spectrum of routing strategies between local system optimal routing and user equilibrium routing and a similar spectrum of control policies can be modeled. These possibilities allow the estimation of the likely consequences of only partial driver and signal compliance with the guidance strategy and the control policy respectively. Plainly, a truly vast spectrum of control/routing possibility is opened up, although not all combinations are of interest.

Modeling Assumptions and Their Relevance

As far as current results are concerned, the road network, including such items as the lane markings and saturation flows, is regarded as fixed; so incidents are not being modeled directly at the moment. All the results here are concerned solely with the minute-to-minute and day-to-day interaction between the control system and the routing strategy, assuming that network characteristics stay the same and any guidance is accurately obeyed. We shall primarily be concerned with this interaction when there is significant or severe recurrent congestion on at least part of the network.

The results involving local system optimal routing will be most relevant if a supportive road pricing system is either in force or being contemplated. But these results are also relevant if guidance is related to an incident or special event on the network which (*a*) renders drivers' knowledge of the network ineffective, and (*b*) creates a more urgent need to maximize effectiveness. Nonetheless in this paper no incident related results are presented, and the road pricing issue is not commented upon. The model would perhaps be useful if the network is changed for a significant period of time by an incident; as it could be applied to this altered network. But at the moment the model does not apply to the transient dynamics which immediately follow the beginning and the end of an incident.

Here the departure times of vehicles are supposed given. The program was extended further to allow for the ability of drivers to choose their departure time in the light of anticipated congestion and their desired time of arrival.

RESULTS

In presenting the results we have chosen to imitate the graphical form of the ordinary single link cost flow function; but here we draw a graph of total travel cost (or time) as the demand level increases. To obtain the graphical results, first a network, a control policy, a dynamic demand, and a (control, routing) pair are chosen. Then, for a proportion p of the fixed demand, the modified CONTRAM model is used to determine the dynamic (traffic control policy, route guidance strategy) pair which satisfies both the route-choice strategy (user equilibrium or local system optimal) and the control policy (fixed-time, equisaturation, P_0, or local delay-minimization at the moment). Then the total travel cost (or time) incurred is calculated for this proportion p. Finally we have increased the demand level p from 0 to 1, in steps of 0.1, and plotted the total cost as a function of p, joining the ten data points by a smooth curve.

In the results presented in the next subsection, the equisaturation policy and the local delay-minimization policy are called equisat and delmin, respectively.

First Realistic Results

These are summarized in Figures 1 and 2. For each demand level there are eight possible total costs according to the (control policy, routing strategy) combination employed. (Some combinations are of little interest.) For an increasing sequence of time-varying demands, the number of vehicle minutes spent for each control/routing combination is shown. The results suggest that on this network local system optimal route-guidance, if totally complied with, together with either of the

TABLE 1 TABLE SHOWING THE (CONTROL, GUIDANCE) PAIRS WHICH CAN BE TESTED BY (*a*) THE ORIGINAL CONTRAM PROGRAM, AND (*b*) THE NEW EXTENDED VERSION OF CONTRAM

CONTROL POLICY	GUIDANCE STRATEGY: User equilibrium	GUIDANCE STRATEGY: Local system optimum
Fixed time	CONTRAM	Extended CONTRAM
Equisaturation	Extended CONTRAM	Extended CONTRAM
Local delay minimisation	Extended CONTRAM	Extended CONTRAM
P_0	Extended CONTRAM	Extended CONTRAM

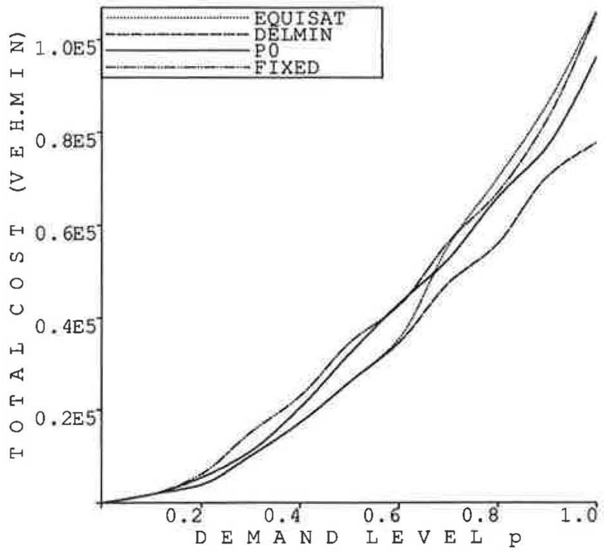


FIGURE 1 Performance with user equilibrium routing and four control policies.

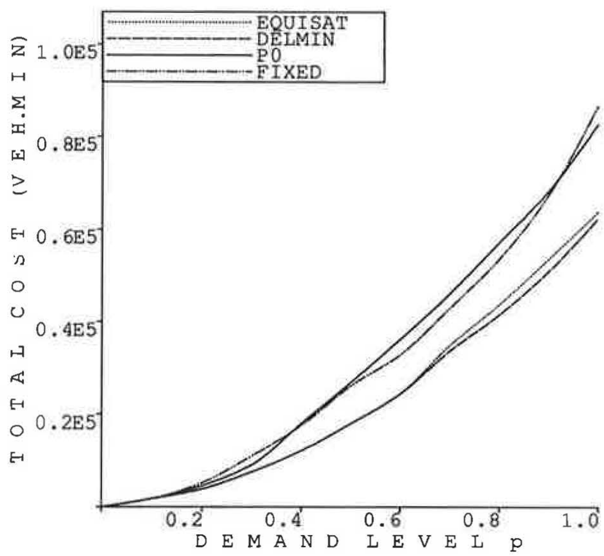


FIGURE 2 Performance with local system optimal routing and four policies.

two familiar policies, may be expected to reduce total travel times by at least 20 percent; and that the control policy (with no route guidance) may be expected to cause the total travel time to vary by about 10 percent. The equisaturation policy is a natural extension of the well-known Webster's method to our current case in which there are queues. It is the existence of these results which holds promise for the future; the actual results presented here are relatively unimportant—they derive from one undoubtedly unrepresentative network. Here only these further observations on this example are made:

1. The performance of the equisaturation policy improves dramatically under local system optimal routing.

2. Local delay minimization is the best control policy with both routing assumptions.

The most important feature of these results is their existence; bearing in mind that they derive from a pretty realistic dynamic model of a realistic congested road network. The network is shown in Figure 3; a much altered part of a town in England; and the stage structure of some junctions is shown in Figure 4.

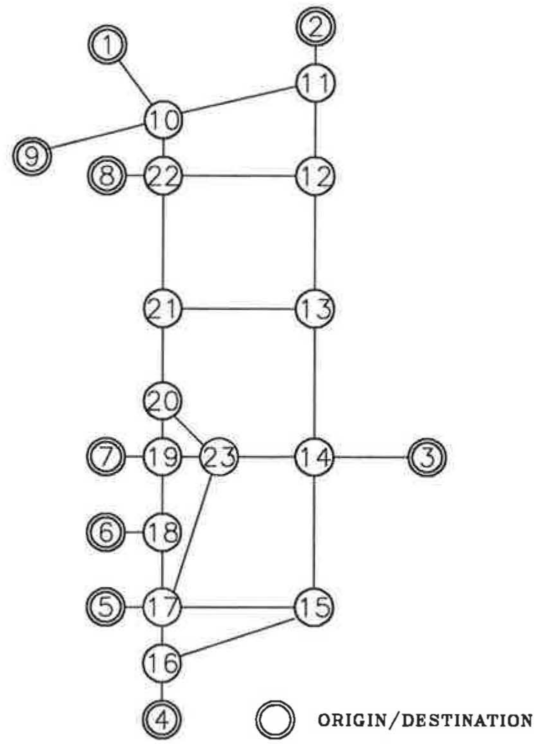


FIGURE 3 Network geometry.

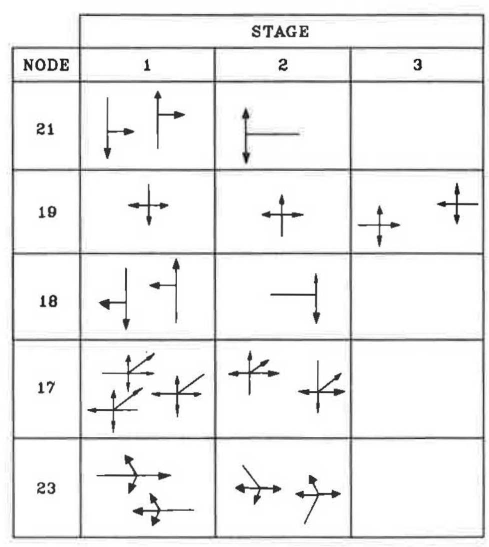


FIGURE 4 A selection of signal stages.

Results Obtained Using an Artificial Network

In these tests only user-equilibrium routing was considered. Figure 5 gives the network geometry: the unlabeled junctions are signal-controlled and those marked F are fly-overs or very large junctions offering, on average, zero or very slight hindrance to traffic flow. Two different choices for the unlabeled, controlled, junctions were considered. In the first, all eight signalized junctions are identical and symmetrical; the results are shown in Figure 6. In the second, each central route is far wider (in fact six times wider), at the signalized junctions, than the outer eight routes. The results here are given in Figure 7. With symmetrical equal junctions, P_0 and the equisaturation policy are identical and so give the same performance, and local delay-minimization is much worse; with the wide central route, P_0 and local delay minimization have similar performances and the equisaturation policy is much worse. The results shown in Figures 6 and 7 suggest that the comparative performances of different control policies are likely to be highly dependent on the detailed structure of the network being considered.

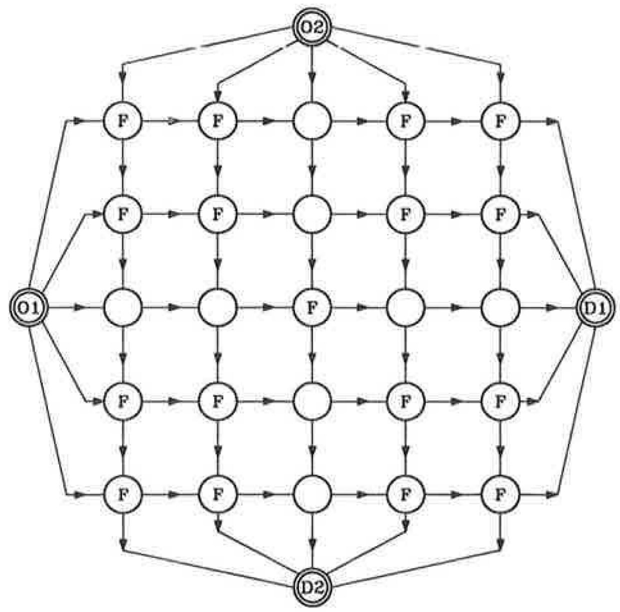


FIGURE 5 An eight signalized-junction artificial network.

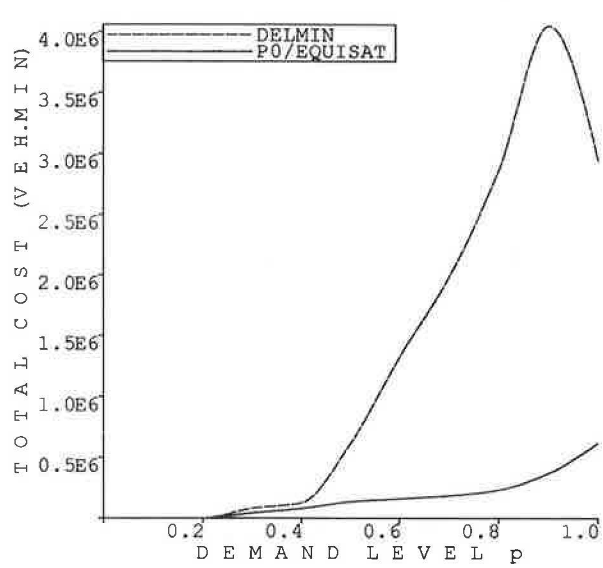


FIGURE 6 Performance of the network when the signalized junctions are symmetrical.

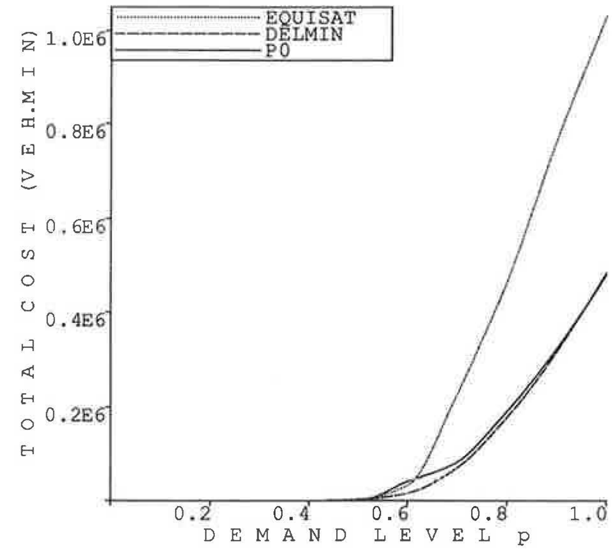


FIGURE 7 Performance of the network when the signalized junctions are unsymmetrical.

CONCLUSION

Performance estimates, summarized in Figures 1 and 2, for all eight combinations of four traffic control policies and two routing strategies (on a realistic network) using our extensions to the CONTRAM dynamic assignment program were obtained. The results suggest that, for this network and a rigid demand, performance gains due to signal control policies are likely to be of the same order of magnitude as performance gains deriving from route guidance (with road-pricing). The results for the artificial network suggest that the best signal control policy will probably depend on the network being considered: judging by these results alone, it is unlikely that a single best control policy exists. The central conclusion is that CONTRAM, together with the extensions described here, is likely to be a very useful tool for assessing the performance of signal-controlled road networks whether route guidance is involved or not. In Figures 1 and 2 each graph represents 10 data points (1 for each demand level); each data point derives from one CONTRAM assignment/control equilibration; and CONTRAM uses 13 time slices. So these eight graphs represent about the same information as that obtained in $10 \times 13 \times 8 = 1{,}040$ static assignments.

REFERENCES

1. H. S. Mahmassani, J. C. Williams, and R. Herman. Performance of Urban Traffic Networks. Proceeding's 10th International Sym-

posium on Transportation and Traffic Theory, Elsevier, 1987, pp. 1–20.
2. R. Herman and I. Prigogine. A Two-Fluid Approach to Town Traffic. *Transportation Science*, Vol. 204, 1979, pp. 148–151.
3. D. R. Leonard, J. B. Tough, and P. J. Bagueley. CONTRAM, a Traffic Assignment Program for Predicting Flows and Queues in Peak Periods. Report LR 841, Transport and Road Research Laboratory, Crowthorne, United Kingdom, 1978.
4. D. R. Leonard, P. Gower, and N. B. Taylor. CONTRAM: Structure of the Model. Research Report RR 178, Transport and Road Research Laboratory, Crowthorne, United Kingdom, 1989.
5. N. B. Taylor. CONTRAM 5: An Enhanced Traffic Assignment Model. Research Report RR 249, Transport and Road Research Laboratory, Crowthorne, United Kingdom, 1990.
6. R. E. Allsop. Some Possibilities For Using Traffic Control to Influence Trip Distribution and Route Choice. Presented at the Sixth International Symposium on Transportation and Traffic Theory, Sydney, Australia, 1974, pp. 345–374.
7. M. J. Smith. A Local Traffic Control Policy Which Automatically Maximises the Overall Travel Capacity of an Urban Road Network. Presented at the International Conference on Urban Traffic Control Systems, University of California, Berkeley, Calif., 1979; Proceedings, ITS Publication No. UCB-ITS-P-79-4; and *Traffic Engineering and Control*, Vol. 21, 1980, pp. 298–302.
8. M. J. Smith. Properties of a Traffic Control Policy Which Ensure the Existence of a Traffic Equilibrium Consistent With the Policy. *Transportation Research*, Vol. 15B, 1981, pp. 453–462.
9. M. J. Smith. A Theoretical Study of Traffic Assignment and Control. Presented at the Eighth International Symposium on Transportation and Traffic Theory, Toronto, Canada, 1981.
10. M. J. Smith. Traffic Assignment and Traffic Control on a Capacity Constrained Network with Queueing. Presented at the Tenth International Symposium on Transportation and Traffic Theory, Boston, Massachusetts, 1987.
11. M. J. Smith, T. van Vuren, D. Van Vliet, and B. G. Heydecker. The Interaction Between Signal Control Policies and Route Choice. Presented at the Tenth International Symposium on Transportation and Traffic Theory, Boston, Massachussetts, 1987.
12. M. J. Smith and M. Ghali. The Dynamics of Traffic Assignment and Traffic Control; A Theoretical Study. Presented at the joint Italian/U.S.A. Seminar on Congested Urban Networks; Traffic Control and Dynamic Equilibrium, Capri, Italy, 1989; *Transportation Research*, Vol. 24B, 1990, pp. 409–422.
13. B. G. Heydecker. Some Consequences of Detailed Junction Modeling in Road Traffic Assignment. *Transportation Science*, Vol. 17, No. 3, 1983, pp. 263–281.
14. M. J. Smith and M. Ghali. Dynamic Traffic Assignment and Dynamic Traffic control. Presented at the Eleventh International Symposium on Transportation and Traffic Theory, 1990.
15. J. Holroyd, Joyce, and J. A. Hillier. The Glasgow Experiment: PLIDENT and After. Report LR 384, Transport and Road Research Laboratory, Crowthorne, United Kingdom, 1972.
16. J. Holroyd and D. I. Robertson. Strategies for Area Traffic Control Systems: Present and Future. Report LR 569, Transport and Road Research Laboratory, Crowthorne, United Kingdom, 1973.
17. P. J. Tarnoff. Concepts and Strategies—Urban Street Systems. Presented at the International Conference on Urban Traffic Control Systems, Berkeley, Calif., 1979; Proceedings ITS publication No. UCB-ITS-P-79-4, Berkeley, Calif., 1979.
18. R. D. Henry, R. A. Ferlis, and J. L. Kay. Evaluation of UTCS control strategies—Executive Summary. Report FHWA-RD-67-149, Federal Highway Administration, Offices of Research and Development, Washington D. C., 1976.
19. D. Van Vliet. SATURN—A Modern Assignment Model. *Traffic Engineering and Control*, Vol. 23, 1982, pp. 578–581.
20. L. Breheret, N. B. Hounsell, and M. McMacdonald. The Simulation of Route Guidance and Traffic Incidents. Presented at the 22nd Annual Universities Studies Group Conference, Hatfield, U. K., 1990.
21. J. C. Smith and K. Russam. Some Possible Effects of Autoguide on Traffic in London. Proceedings of First Vehicle Navigation and Information Systems Conference, Toronto, Canada, 1989, pp. 282–287.
22. J. A. Charlesworth. The calculation of mutually consistent signal-settings and traffic assignment for signal controlled road network. Presented at the Seventh International Symposium on Transportation and Traffic Theory, Kyoto, Japan, 1977.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

Special-Purpose Parallel Computer for Traffic Simulation

H. J. M. VAN GROL AND A. F. BAKKER

Traffic simulations are widely used for long- and medium-term forecasting of traffic. Now-a-days, with the growing problem of queues during rush hours, the demand arises for dynamic traffic management and, therefore, short-term forecasting. Apart from the need for new, adapted dynamic assignment models the second important part in this new development is the required computational power. Most commercially available computers are unable to produce an accurate forecasting for the next 5 to 30 min. within the desired time and budget. Analysis of existing assignment models and their most time consuming part—shortest path finding—has shown that the main structure of the models can be parallelized. The use of parallelism thus seems apparent. Several general purpose parallel computers, such as N-cubes, are commercially available. However, apart from being more expensive, they loose a large part of their expected performance by the amount of necessary interprocessor communications. Additionally, the programming of such computers has turned out to be more difficult than expected. A simple linear array of typically 16 processors, the so called Linear Processor Array (LPA) is proposed. This one-dimensional parallel computer with high-speed buffered interconnections between each pair of neighboring processor boards, parallel accessible by both a control board and a general-purpose host computer, forms a transparent concept for the programmer. The optimally configured boards together with the high speed intercommunication allow a cost/performance improvement factor of 100 compared with a minisupercomputer like the Convex C1. LPA should be a powerful tool for future developments of on-line traffic control, route guidance, ramp metering, among other things.

Many transportation planning and control activities involve traffic simulations for medium- and long-term forecasting. The existing models are efficient enough to do this within the desired time. With larger networks, the simulation time increases rapidly. Furthermore, with the growing importance of dynamic traffic management (because building new roads is considered not to be a desired solution anymore), the demand for short-term forecasting grows. Existing models do not meet these demands, neither in time nor in accuracy.

Apart from the development of new adapted dynamic assignment models, the time constraints require thought about the necessary computing power. With most commercially available computers it is not possible to produce an accurate forecasting for the next 5 to 30 min within the desired time and budget. The use of special-purpose computers will open ways to a desired solution.

A cost-effective answer to this problem is given in this paper. To start with, the existing assignment models and the development of the new dynamic models will be investigated. This will lead to a simplified representation of assignment models in which the algorithm is analyzed to find possible optimizations. It will be shown that the main structure of the models can be parallelized. After a short discussion about the use of general-purpose computers and an introduction to special-purpose computers, the use of special-purpose hardware for traffic routing problems will be discussed. This will lead to the proposed Linear Processor Array (LPA), which will be explained in detail. Finally, some preliminary results will be presented and the expected performance improvement will be given in conclusion.

Delft University of Technology, Faculty of Applied Physics, Physics Informatics/Computational Physics, Lorentzweg 1, 2628 CJ Delft, The Netherlands.

TRAFFIC SIMULATION MODELS

Traffic simulation models or, more specifically, network assignment models have evolved from the simple, static assignment models (all-or-nothing assignment) to the more complex dynamic assignment models (three dimensional assignment model). In the past, the models have mainly been used for long- and medium-term forecasting and have played an important role in transportation development schemes. Long computations are manageable in these cases. The computation time, however, will grow rapidly with more complex models, larger networks and so on. A third important objective, short-term forecasting, has recently gained attention. Short-term forecasting is used for dynamic traffic management, which deals with on-line networkwide traffic control, route guidance, and ramp metering, for example. Short-term forecasting, in the range of 5 to 30 min, will impose a time constraint on the simulations. All together there exist enough reasons to justify a search for methods to speed up the calculations. Improvements can be made on both the algorithm side and the computer side. In the remainder of this paper possible improvements on the algorithm side will be concentrated on.

First, the collection of assignment models will be looked at to find a generalized form. Although the reader will be familiar with the models, they will be summarized.

Categorization of the Models

The simplest model is the all-or-nothing model. This model, however, does not take into account the multiple routes different travelers, from one origin-destination (O-D) pair, take in reality; nor does it take into account the load dependence of travel times or time dependence in general. The remaining models can be split into three categories, on the basis of their resolution of these three deficiencies. The corresponding algorithms are, respectively, the stochastic, the equilibrium, and the dynamic assignment algorithms. A few examples of each category will now be listed.

Stochastic Assignment

The principle of stochastic assignment is to simulate the route choice by assuming the traveler does not have perfect knowledge about his route. There are two essentially different stochastic assignment methods. The first one iteratively performs all-or-nothing assignments on the basis of randomly adjusted arc travel times. The assignment at each iteration is a fraction (1/number of iterations) of the total assignment. The maximum number of possible routes between one origin and destination is equal to the number of iterations. The second method consists of only one or two iterations but results in a large number of taken routes. Both methods have been described in detail in Van Vliet (*1*). These methods are mainly used in noncongested networks.

Equilibrium Assignment

In this category, the travel times depend on the traffic load. Multiple routes are found at equilibrium. The equilibrium is based on the principle of Wardrop (*2*): "The traveltimes of all used routes between an origin and a destination are equal to or smaller than the traveltimes of the unused routes. No traveler can shorten his journey by switching to another route." There are several ways to reach the equilibrium: methods with a theoretical background and more ad hoc methods. The practical difference is the number of iterations necessary to reach the equilibrium and the amount of work in each iteration. At the end of each iteration travel times are recalculated. Methods include

- Equilibrium: the equilibrium assignment is computed using optimization techniques. At each iteration the old assignment is compared with a new all-or-nothing assignment. The resulting assignment is the optimal weighting between these assignments. This is done until no significant improvement can be made.
- Capacity restraint: the network is iteratively loaded and unloaded, according to a predefined recipe.
- Fixed demand incremental: the network is iteratively loaded with a decreasing amount $1/(i + 1)$, where i is the iteration number.

Dynamic Assignment

Instead of observing the assignment in one time interval, assuming the total amount of traffic on a route to be evenly spread along the route, time aspects are taken into account. In other words, by dividing the observed interval into a number of periods, travelers are only contributing to the load of an arc in the period in which they are really using it. Since travelers do not start their journey all at once, the O-D matrix is defined for each period separately.

One dynamic assignment method is described by Hamerslag (*3,4*). For each period an all-or-nothing assignment is computed, resulting in an all-or-nothing assignment in three dimensions (time and space). Techniques from the previous categories are used to find a realistic assignment. The number of iterations, required in the former algorithms, is now multiplied by the number of simulated periods. This method can realistically simulate traffic in congested networks. It is possible to simulate temporary decreasing capacities caused by, for instance, accidents or ramp metering.

Summary

It can be concluded that all algorithms are repetitive all-or-nothing assignments. The workload (computational load) comes almost entirely down to the all-or-nothing assignment. To speed up the computation, optimizing the all-or-nothing assignment must be concentrated on.

The All-Or-Nothing Assignment

The all-or-nothing assignment, as the name suggests, assigns "all" traffic to a single shortest route and "non" to the others. It will, therefore, compute the shortest path between each O-D pair and assign the associated amount of traffic to each consecutive arc along that path.

```
# Compute the all-or-nothing assignment
for each O-D pair in the network {

    Compute the shortest path between origin and destination

    Assign the path

}
```

It is more efficient to calculate shortest path trees (*spt's*), because the corresponding parts of the shortest paths from one origin to several destinations can be long. Thus

```
# Compute the all-or-nothing assignment

for each origin in the network {

    Compute the spt

    Assign the spt

}
```

We will now examine the algorithm in more detail.

Shortest Path Finding Algorithms

There are four main techniques of finding shortest paths:

- Heuristic technique—one-to-one;
- Algebraic technique—all-to-all;

- Combinatorial technique—one-to-all; and
- Optimizing technique—one-to-all.

For this study, the heuristic technique is not suitable (one-to-one) and the algebraic technique is inefficient and uses much memory. This leaves the combinatorial and the reoptimizing techniques. The reoptimizing technique is the fastest technique and is based on the reoptimization of an existing *spt* for origin u to a new *spt* for origin v. As it is a more complex algorithm, consumes a large amount of memory, and is only about two times faster, one might prefer to use the combinatorial technique.

The algorithms using the combinatorial technique can be split into two groups, the label-correcting and the label-setting algorithms. The labels are the values associated with each node in the network representing the cost of traveling coming from the present origin-node (root of the *spt*).

The following algorithm is a general shortest path finding algorithm:

```
# Compute the spt

Initialize  d_v = ∞, v ∈ N − {r}, d_r = 0

Init_Q  Q = {r}

While (Q ≠ φ) {

  Select_node select u ∈ Q; Q = Q − {u}

  for each (u,v) ∈ FS(u) {

    if (d_u + c_u,v < d_v) {

      p_v = u

      d_v = d_u + c_u,v

      Update_Q  Q = Q ∪ {v}

    }

  }

}
```

For an explanation of the symbols used, see Appendix.

The tree is built from the root; therefore, these algorithms are also called tree-builder algorithms. The set Q contains the nodes that need to be examined. All algorithms are similar except for the way Q is maintained in **Init_Q**, **Select_node**, and **Update_Q**. The efficiency of the algorithm depends on the way this is done. The minimum computation time is reached when **Init_Q**, **Select_node**, and **Update_Q** cost minimal time.

Well-known algorithms are Moore (*5*), which is a label-correcting algorithm and Dijkstra (*6*), which is a typical label-setting algorithm. A more detailed description can be found in (*7–10*). A simple label-setting algorithm is one that simply sorts the entries in Q, (S-ord). It has a time complexity of $O(m.n)$, where m is the number of arcs in the network and n the number of nodes. The fastest algorithm using the combinatorial technique is a threshold algorithm (T-calc), which combines the good qualities of both the labeling methods. Although, in principle, the computational complexity of this algorithm is $O(n.2^n)$, and thus about the worst possible, in practice it behaves like $O(n)$ and is robust. For the maintenance of Q it uses a combination of methods: address calculation and a lifo-procedure (last-in-first-out) [see van Grol & Bakker (*11*)]. In this way, the algorithm dependent parts (**Init_Q**, **Select_node** and **Update_Q**) cost minimal time. The differences between the methods increase with growing network size. Having reduced the time complexity of the *spt* to $O(n)$ the time complexity of the assignment part will now be examined.

Assignment

The simplest way of assignment is to follow the O-D path and assign the associated amount. The following is the procedure to assign one shortest path tree (q_a, $a \in M$ contains previously assigned loads):

```
# Assign the spt for origin

for each destination in the network {

  u = destination

  while u is not equal to origin {

    q_pu,u = q_pu,u + OD(origin, destination)

    u = p_u

  }

}
```

In this way the arcs near to the origin are assigned many times. The time complexity is $O(n.l_r) = O(n.n)$, where l_r is the mean number of arcs in a path.

A more efficient method can be obtained by simultaneously assigning several O-D elements to an arc. Supposing that l_u is the level in the tree (the number of nodes counted from the origin), all nodes can be sorted according to their level. Starting at the highest level, the following procedure can be executed.

```
# Assign the spt for origin

K = {1 . . . n}, S_u = 0, u ∈ N

# Sort the nodes in set K top-level down

Sort K

for each node u from K {

  q_pu,u = q_pu,u + OD(origin,u) + S_u

  S_pu = S_pu + OD(origin,u) + S_u

}
```

The set K contains all nodes with their levels. This reduces the complexity of the assignment to $O(n)$ but sorting K has a complexity of $O(n.\log n)$. By using an addressable array, $L()$, to sort the nodes by their level we can reduce the total complexity to $O(n)$. The algorithm then becomes

Sort levels of *spt* for origin

$L(.) = 0$

for each node u {

$L(l_u) = L(l_u) \cup \{u\}$

}

Assign the *spt* for origin

$S_u = 0, u \in N$

for each level top down {

for each node u in the set L(level){

$q_{p_u,u} = q_{p_u,u} + OD(\text{origin},u) + S_u$

$S_{p_u} = S_{p_u} + OD(\text{origin},u) + S_u$

}

}

Summary

A complete description of the algorithm can be found in the Appendix. The overall computational complexity (path finding and assignment) for one *spt* is now $O(n)$. This means that a number of operations is executed on each node and $O(n)$ is thus the minimum complexity. Only the number of operations can now be minimized. The computational complexity of the all-or-nothing assignment is $O(n^2)$. A good implementation of this algorithm is the best that can be done. A major improvement could only be achieved by an implementation in assembler.

Two observations can be made. First, it can be seen that the computations of the *spt*'s are independent. This allows the use of parallelism. The second observation is that the all-or-nothing assignment is mainly a data flow problem. All network data will flow through the algorithm several times, whereas the number of operations on the data is minimal. Bearing this in mind, the possible use of special-purpose hardware is discussed in the next section.

SPECIAL-PURPOSE HARDWARE

First, the use of general-purpose computers in traffic simulations will be discussed, because these limitations motivated the current research into using special-purpose hardware. The principles of the special-purpose computers are described next. Third, dedicated architectures for traffic simulations will be focused on. After an introduction to the proposed LPA, a more detailed description of the LPA, dedicated for traffic simulations, will conclude the section.

General-Purpose Computers

Standard traffic simulations usually run on general-purpose computers, such as microcomputers, workstations, and mainframes. For small networks and simple algorithms, the turnaround time of the simulations is satisfactory for most applications. Moving from microcomputer to mainframe or minisuper workstations significantly improves performance and visualization, but larger networks and more complex assignment algorithms require supercomputer power. However, the cost and the limited availability of supercomputers eliminate this option.

In general, commercially available computers were designed to solve all problems, and are not tailored to efficiently solve a typical problem such as found in traffic simulations. To improve the cost to performance ration, or to bypass hardware limitations of general-purpose computers, one can design and build a special computer, the architecture of which maps perfectly on the problem or algorithm involved. This approach can be considered a low-cost alternative to supercomputers.

Special-Purpose Computers

Special-purpose computers are designed to efficiently carry out a particular task at supercomputer speed. In general, they cannot handle any other task, or, if they can, the performance will be poor. However, a design can cover a class of problems, and thus can be used for a wide range of applications without performance penalty. The special-purpose computer is designed on the basis of the problem, problems, algorithm, or algorithms to be used, and allows an architecture that explores parallel and pipelined operations wherever applicable. It allows problem-dependent memory organization, problem-adapted basic instruction set, and so on with the purpose to improve the total performance of the computer.

Special-purpose computers range from single-purpose to multi-purpose computers and they differ in the flexibility of programming them.

- *Single-purpose computers.* Single-purpose computers have a basic instruction set that will only cover the operations required for the task it has been designed for. This approach allows the algorithm to be hardwired, which guarantees an optimal speed. Fixed-wired parallel and pipelined architectures restrict the flexibility to modify the algorithm, but leave open the possibility to vary enough parameters to motivate the effort to design such a machine. The cost to performance ratio of this type of computer is low (factor 100 better than supercomputers) and they are available 24 hr/day for the computer experimentalist. A variety of single-purpose computers have been successfully exploited in signal processing and computational physics (*11–14*).
- *Multi-purpose computers.* Multi-purpose computers are designed to efficiently solve a class of problems rather than a single problem. These architectures reflect the common

property of the algorithms involved and can be programmed to solve a problem from the class of problems it was designed for. High-level languages are used (C and F77) to program the computer. The speed is obtained by using many processor nodes interconnected by a communication network that is suitable for the class of algorithms involved. The architecture of the nodes is kept simple but effective to allow the construction of efficient compilers. The choice of processor, memory structure and size, interconnection network, and word length characterize the multi-purpose computer. They can be shared memory or distributed memory machines running in single-instruction multiple-data or multiple-instruction multiple-data mode. Flexibility and programmability of these computers are traded for ultra speed as in single-purpose computers. However, the newest commercially available microprocessors are fast, the architecture is scalable, and can result in a cost to performance ratio improvement by two orders of magnitude compared with supercomputers.

So-called general-purpose parallel computers, such as N-cubes, may use fast processor nodes, but their memory architectures and their slow interconnection networks do decrease the overall performance dramatically. Only between 10 and 20 percent of the advertised peak performance is reached by careful programming. Existing parallelizing high-level language compilers are still far from ideal. Automatic decomposition of sequential program flows of problems that are often parallel in nature is not an efficient way to obtain fast codes for parallel computers.

In practice, the user has to choose a network topology and program the nodes to use that network efficiently. Often topologies are chosen to be ring structures, so that the programmer can implement the algorithm without begin distracted by more exotic topologies. Still, the node interconnections are slow because of their all topology structure.

A Special-Purpose Computer For Traffic Simulations

To select or design a computer for traffic simulations, the algorithm has to be examined for possible parallel or pipelined operations. As the algorithms involved are still in development, thus demanding flexibility, a single-purpose computer will not be considered. Decomposition of the total problem into coupled parallel processes is a way to find a scalable parallel architecture. The node architecture, memory size, and interconnection channel will then determine the final computer.

As shown previously, a time critical part of the algorithms used in traffic simulations is the calculation of n shortest path trees (*spt*'s), where n is the number of nodes in the network. It was concluded that the *spt*'s can be independently calculated, thus allowing us to decompose the problem into n problems of one *spt*. Using n processor boards, all the *spt*'s can be calculated in parallel. This will improve the simulation speed by order n. In a parallel computer, with P processor boards, n/P *spt*'s can be calculated on each board, which gives a speed-up factor of P. The latter solution is preferable for reasons of scalability, especially for large n. Each processor board will need all network information, thus P times the amount of memory needed to store the network is the minimal total memory size. Minimizing P to keep memory costs down is compensated by making the processor boards as fast as possible, and keeps the overall performance high. In the preceding assignment schemes, this decomposition also holds. However, to find the total load per arc, the partial arc loads must be accumulated. It is necessary, therefore, to efficiently interconnect the processor boards. Here, a pipeline structure in which each processor board is one pipe stage of the whole pipe can be used. Consequently, the processor boards are ordered in one string to construct this pipeline. One high-speed data channel from each board to its neighbor board is sufficient to obtain a fast pipeline. The accumulated arc loads are collected in the last pipe stage (last processor board in the string). To start the next iteration, the updated network has to be broadcast to all processor boards. When all processors are connected to a common bus, broadcasting can be accomplished using one talker and P listeners on this bus (see Figure 1). Before going into details of the design, a more generalized architecture that will cover the above architecture ideas for traffic routing simulations but can be used for other purposes as well (multi-purpose computer) is discussed.

LPA Architecture

An LPA is a one-dimensional array of identical processor boards, each of which is connected to its two neighboring boards only by a data bus. In addition, the boards share a common data-, address-, and control-bus, which is also interfaced to a general-purpose host computer. One processor board is configured as a control board, which supervises the chain of processor boards through a special control bus (see Figure 2).

A large class of problems in computational physics (both authors work in this field) can be solved using this parallel architecture. A natural domain decomposition allows the mapping of different subdomains on different processors. All subdomains can be processed simultaneously. In general, the calculations in a subdomain need data from the other subdomains, but when "local environment problems" are

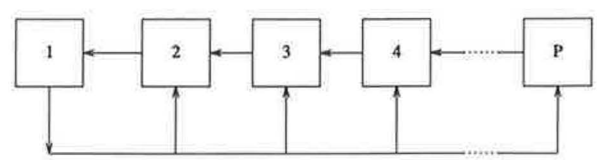


FIGURE 1 Data flow in the traffic assignment problem mapped on an array of P processor boards.

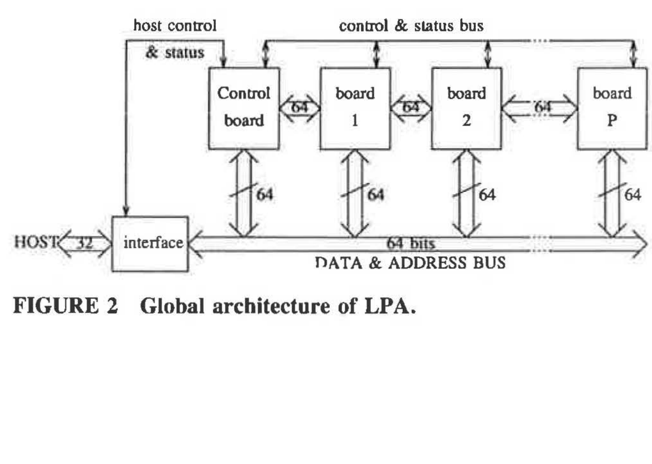


FIGURE 2 Global architecture of LPA.

involved, such as in finite difference calculations, only the directly neighboring subdomains are contributing to the results. For this class of problems, the domain can be decomposed by slicing in just one direction. Each domain slice is mapped onto one processor, and each processor communicates only with its two neighboring processors. The host passes data and programs to the LPA and collects results of the calculations by means of the common bus. The processor architecture, the local memory and the interface to the two neighboring processor boards are optimized to tackle the problem they are designed for.

LPA for Traffic Simulations

LPA architecture is well suited for the traffic-routing problem. Programs and network data, which are (mostly) identical on the different boards, can be broadcasted to the processor boards. Each processor node requires enough memory to contain all network information and the locally calculated results. At the start of every iteration (all-or-nothing assignment), the network information on all the processor nodes is identical. Then, every node starts to calculate its share of the total number of shortest paths instructed by the control board. Clearly, this calculation is intensive. A fast processor is needed for all floating point calculations. In addition, because a lot of data are involved, the memory interface is important. Bus contention should be avoided at all times. This is partly solved by the parallel approach taken, with independent nodes and distributed memory, but a fast memory interface tuned for this particular problem is still called for.

The processor needs to be selected primarily on grounds of floating point calculation speed. The fastest processor available today is the Intel i860 Microprocessor. It contains a core unit, floating point unit, and instruction and data caches on one chip. Theoretical speeds are 40 million instructions per sec (MIPS) and 80 million floating-point operations per sec (MFLOPs). Practical speeds using high-level languages such as C and Fortran are in the order of 15 MFLOPS. Hand coded assembly is capable of performing between 30 and 80 MFLOPS, depending on the algorithm, and the amount of vectorization and pipelining that can be employed. Furthermore, the Intel i860 can execute integer and floating point operations in parallel and has a 64-bit-wide memory bus.

Because of the speed at which the processor processes data, the memory interface has become the only bottleneck. The rate at which data can be retrieved from and stored into memory determines the overall processing speed. The network information is typically scattered through memory, thus rendering the on-chip (small) data cache practically useless. Therefore, the interface to the dynamic memory (use of static memory only would be too expensive) is vital.

At the end of each iteration, every board has calculated part of the load for every arc. These partial loads can be accumulated in a pipelined fashion using the connections to the neighbor boards: every board receives the partial loads from its right neighbor, adds its own partial loads, and hands the results to its left neighbor. Finally, the accumulated partial loads are handed to the control board (which is the left-most board), which can start the next iteration by broadcasting the updated network.

The connections between the boards are realized by first-in first-out buffers (FIFOs), which are 64-bit-wide memory components that move data in receiving order to the neighbor board (size of buffer is several kByte). The FIFOs are used to automatically synchronize the asynchronous processors, and buffer data sent between them. This allows the processor nodes to act autonomously and send data whenever ready. The control board is not needed to synchronize nodes or buffer data. All processor nodes can use their FIFOs concurrently, allowing for maximum throughput.

RESULTS

It was previously shown that the performance of traffic simulation programs can be improved and how special-purpose hardware can be used to reduce the computing time and the cost involved. The improvements expected by using special-purpose hardware will now be defined.

First, the improvement by parallelization will be considered. Using an architecture by the LPA concept is proposed, with 16 i860-based processor boards, an improvement of a factor 16 compared with one i860-based processor board can be expected. Although this seems obvious, most computers with parallel processors are unable to improve their performance by the number of processors they use—compared with one processor [see (*15,16*)]. The improvement is justified by the negligible overhead in interboard communication. The interboard communication, required to accumulate the arc loads and to broadcast network data, does not increase with the number of processor boards and is small compared to the total computation.

An operating system running on a computer allows the users to use all kinds of facilities, such as file-support, I/O in general, multitasking, scheduling, timing and so on. Using such an operating system will decrease the overall performance of the system. By avoiding most of these facilities some performance improvement can be gained. The operating systems running on the LPA nodes will allow the minimum amount of facilities to run the problem efficiently.

Second, the expected performance of the LPA with some general-purpose computers that are commonly used are compared. The i860 is a fast processor as explained previously. The processor alone is already in competition with several fast general-purpose computers. Some test runs have been executed on the Intel Microprocessor Software Development system—STAR860—which is an AT-386 with an i860 CPU based add-on card. This board is not optimally configured, and can thus be improved on the performance in the final design.

Comparison tests have been run on a Convex C1, several differently configured Silicon Graphics (SG, R3000, 33-25-20 MHz, 64k-Byte cache), a MicroVAXIII, a MicroDutch (68020), an Hewlett Packard Workstation (HP, 68030), and a personal computer (AT-386, 25Mhz, 64kByte cache). The Convex C1 is a vector processor with an architecture resembling a Cray supercomputer. Although not the fastest, it is a widely used computer. The test programs were written in C. Two shortest path tree algorithms discussed previously, S-ord and T-calc, were used. The assignment was done in two different ways, the 'simple' assignment and using levels, also described previously. For the implementation of the algorithms we can use pointers or indexing. The calculations can be done in integers or in floating point notation. The resulting computing times are given in Tables 1–3.

On all computers available optimizers were used. The network used contained 3,347 nodes, 9,394 arcs, and an arc to node ratio 2.8. As comparison, a network of 17,931 nodes was also used. The computing times given are the all-to-all times; an all-or-nothing assignment with each node as origin and destination. The i860 is not the fastest processor in Table 1 because of its memory configuration. The Silicon Graphics (33MHz) has an advantage of having a 64kByte cache. This advantage disappears when a larger network is used (see Table 3). With integer indexing the i860 is always superior. Next to an upgrade of the i860 from 33 MHz to 40 MHz, the processor board can be improved by using a pipelined-multibank memory instead of the single-bank memory implementation on the STAR860.

CONCLUSIONS

The use of special-purpose hardware is only legitimate when the task to be performed is time critical and the budget is limited. It was shown that the STAR860 is two times as fast as the Convex C1. The performance of a single i860-based processor board, in comparison to the STAR860, can be improved by a factor of about 2, using a faster version of the i860 and a better memory architecture.

The price of a single i860-based processor board is mainly determined by the memory cost and, depending on the amount of memory, is estimated to be between $5,000 and $10,000. A 16-board LPA together with a general purpose host, of about $40,000, and additional costs of manufacturing of about $15,000 amounts to a total cost of less than $0.3 million. The price of a Convex C1, however, is about $0.6 million, and thus leads to another factor 2 in cost to performance improvement. Hence, a 16-board LPA will give a cost to performance improvement of a factor 100 compared to a Convex C1. For future developments of on-line traffic control, route-guidance, ramp metering and so on, the LPA should be a powerful tool.

TABLE 2 THE DIFFERENCE IN COMPUTING TIMES BETWEEN USING A SIMPLE ASSIGNMENT METHOD AND ONE USING LEVELS

Data-struct:	Pointer			
spt-alg	T-calc		S-ord	
Assignment	Levels	No-levels	Levels	No-levels
SG (20 MHz)	266	703	315	751
SG (25 MHz)	260	665	302	701
SG (33 MHz)	161	438	189	464
STAR860	190	535	202	548

NOTE: The times are given for two *spt* algorithms and on several computers. The network size N = 3,347. SG is Silicon Graphics.

TABLE 3 THE DIFFERENCE IN PERFORMANCE OF THE COMPUTERS ON NETWORKS WITH DIFFERENT NETWORK SIZES

Size	3,347	17,931
Convex C1	411	683
SG (25 MHz)	260	581
SG (33 MHz)	161	417
STAR860	190	319

NOTE: The times given are in case of N = 3,347, all-to-all, in case of N = 17,931, 1,000 *spt*'s. The *spt*-algorithm used was T.calc and the assignment uses levels. SG is Silicon Graphics.

ALL-OR-NOTHING ASSIGNMENT

The following is a description of the algorithm used to perform the all-or-nothing assignment. This is an optimal algorithm depending on the way **Init_Q**, **Select_node**, and **Update_Q** are implemented. The symbols used are defined as follows:

N, M = set of all nodes, arcs in the network;
n, m = number of nodes, arcs in the network;
$FS(v)$ = the foreward star representation of node v, defines the network;
$OD(r,v)$ = matrix, number of travelers going from node r to v;

TABLE 1 COMPUTING TIMES IN SECONDS ON A NUMBER OF COMPUTERS

Data-struct:	Pointers				Indexing			
Notation:	Floating Point		Integer		Floating Point		Integer	
spt-alg	T-calc	S-ord	T-calc	S-ord	T-calc	S-ord	T-calc	S-ord
Microdutch	4,307	6,155	2,626	2,989	6,372	10,284	4,352	6,966
AT-386	1,485	2,137	733	674	1,807	2,823	840	1,305
MicroVAXIII	1,180	1,712	1,119	1,324	1,830	3,004	1,635	2,585
HP	1,149	1,562	789	764	1,674	2,683	1,133	1,747
Convex C1	411	535	315	382	515	744	384	565
SG (20 MHz)	266	315	237	273	268	365	230	341
SG (25 MHz)	260	302	238	270	319	407	276	393
SG (33 MHz)	161	189	139	164	222	286	187	268
STAR860	190	202	185	192	235	283	179	240

NOTE: The program, either T-calc or S-ord, is implemented with pointers or indexing and the calculations in either integer or floating point notation. The assignment is implemented with the use of levels. The network size N = 3,347. HP is Hewlett Packard, SG is Silicon Graphics.

c_a, $c_{u,v}$ = used arc length (arc travel time) for arc a from node u to node v;
d_v = calculated distance (travel time) from the origin node to node v;
p_v = previous node from node v in the shortest path tree (*spt*), defines the *spt*;
l_v = level of node v in the shortest path tree;
q_a, $q_{u,v}$ = calculated load on the arc a, from node u to node v;
r = current origin-node (root);
Q = temporary set, contains nodes to be examined;
$L(n)$ = set containing all nodes from *spt* on level n; and
S_v = temporary, traffic load going up to node v.

In short, the all-or-nothing assignment looks as follows:

```
# Compute the all-or-nothing assignment

for each origin in the network {

    Compute the shortest path tree

    Sort levels

    Assign the spt

}
```

The subroutines **Init_Q**, **Select_node**, and **Update_Q** are not defined here. Although crucial to the efficiency of the algorithm, the functionality remains the same. The calculated loads from preceding assignments are kept in q_a, $a \in M$.

Compute the shortest path tree for origin r

Initialize $d_v = \infty$, $v \in N - \{r\}$, $d_r = 0$, $l_r = 1$

Init_Q $Q = \{r\}$

While $(Q \neq \phi)$ {

  Select_node select $u \in Q$; $Q = Q - \{u\}$

  for each $(u,v) \in FS(u)$ {

    if $(d_u + c_{u,v} < d_v)$ {

      $p_v = u$

      $l_v = l_u + 1$

      $d_v = d_u + c_{u,v}$

      Update_Q $Q = Q \cup \{v\}$

    }

  }

}

Sort levels

Initialize $L(.) = 0$

for each node u {

  $L(l_u) = L(l_u) \cup \{u\}$

}

Assign the shortest path tree

Initialize $S_u = 0$, $u \in N$

for each *level* top down {

  for each node u in the set $L(level)$ {

    $q_{pu,u} = q_{pu,u} + \mathrm{OD(r,u)} + S_u$

    $S_{p_u} = S_{p_u} + \mathrm{OD(r,u)} + S_u$

  }

}

REFERENCES

1. D. Van Vliet. Road Assignment I, II en III, *Transportation Research*, Volume 10, 1976, pp. 137–157.
2. J. G. Wardrop. Some Theoretical Aspects of Road Traffic Research. *Proc. of the Institute of Civil Engineers*, Part II, 1, 1952, pp. 325–581.
3. R. Hamerslag and P. C. H. Opstal. *A Three-Dimensional Assignment Method in Time-space.* Report 87-49 from the Faculty of Mathematics and Informatics, Delft University of Technology, The Netherlands, 1987.
4. R. Hamerslag. Dynamic Assignment in the Three Dimensional Timespace. In *Transportation Research Record 1220*, 1989.
5. E. F. Moore. The Shortest Path Through a Maze. *Proc. of International Symposium on the Theory of Switching*, Harvard University Press, Cambridge, MA, 1959, pp. 285–292.
6. E. W. Dijkstra. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1, 1959, pp. 269–271.
7. R. Dial, F. Glover, D. Karney, and D. Klingman. A Computational Analysis of Alternative Algorithms and Labeling Techniques for Finding Shortest Path Trees, *Networks 9*, 1979, pp. 215–248.
8. G. Gallo and S. Pallottino. Shortest Path Methods: A Unifying Approach. *Mathematical Programming Study*, 26, 1986, pp. 38–64.
9. F. Glover et al. *Threshold Assignment Algorithm.* Report CBDA 107, Center for Business Decision Analysis, University of Texas, Austin, 1982.
10. H. J. M. van Grol and A. F. Bakker. Shortest Path Finding: From algorithm to Special Purpose Hardware. *Proc., Transportation Planning 17th PTRC Transport & Planning Summer Annual Meeting*, University of Sussex, England, 1989.
11. A. F. Bakker and C. Bruin. Design and Implementation of the Delft Molecular-Dynamics Processor. In *Special Purpose Computers* (Berni J. Alder, ed.), 1988, Chapter 6. ISBN 0-12-049260-1
12. D. J. Auerbach, A. F. Bakker, T. C. Chen, A. A. Munshi, and W. J. Paul. A Highly Parallel Computer for Molecular Dynamics Simulations. *Materials Research Society Symposium Proc.*, Vol. 63, 1985.

13. D. J. Auerbach, W. J. Paul, A. F. Bakker, C. Lutz, W. E. Rudge, and Farid F. Abraham. A Special Purpose Parallel Computer for Molecular Dynamics: Motivation, Design, Implementation, and Application. *Journal of Physical Chemistry*, Vol. 91, 4881, 1987.
14. A. F. Bakker, G. H. Gilmer, M. H. Grabow, and K. Thompson. A Special Purpose Computer for Molecular Dynamics Calculations. *Journal of Computational Physics*, Vol. 90, No. 2, 1990.
15. Huey-Kuo Chen and David E. Boyce, Code Optimization For A Nonlinear Transportation Network Model: A Case Study, *Presented at the 68th Annual Meeting of the Transportation Research Board*, Washington, D.C., 1989.
16. K. C. Mouskos and Hani S. Mahmassani. Guidelines and Computational Results for Vector Processing of Network Assignment Codes on Supercomputers. *Presented at the 68th Annual Meeting of the Transportation Research Board*, Washington, D.C., 1989.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

Dynamic Network Traffic Assignment and Route Guidance Via Feedback Regulation

MARKOS PAPAGEORGIOU AND ALBERT MESSMER

A deterministic, macroscopic modeling framework for dynamic traffic phenomena on networks consisting of freeways and urban streets is presented for nonelastic but time-varying traffic demands. A feedback methodology is applied to the network model to establish dynamic traffic assignment conditions. Specifically, a multivariable feedback regulator with integral parts and a simple bang-bang controller are developed and tested for a particular network traffic model. Because of three fundamental features (low computational effort, low sensitivity with respect to unknown demands and compliance rates, and integrated design procedure), the feedback concept appears attractive for a broad class of traffic control problems including route guidance systems.

Dynamic modeling and control of a multidestination traffic network is generally considered to be a highly complex problem. There is no generally applicable macroscopic mathematical model describing dynamic phenomena of traffic flow on street or freeway networks to the best of the authors' knowledge. Nevertheless, traffic network models are urgently needed both as simulation tools and as a basis for developing efficient route guidance strategies. Perhaps the most advanced concept so far for macroscopic dynamic modeling of multidestination networks is the one proposed by D'Ans and Gazis (*1*). In their work, however, D'Ans and Gazis assume that the route choice of drivers with a given origin and destination is fixed and known. In this paper, dynamic modeling and control of traffic networks including traffic assignment are considered. A basic assumption is that traffic demand at the origins of the network is considered to be deterministic and independent of the traffic conditions in the network. Consideration of elastic demands is left to future investigations.

The model presented in this paper was developed on the basis of a dynamic traffic network model framework that was initially presented elsewhere (*2*). The model consists of three interacting parts:

1. A traffic flow part describing traffic flow evolution along network links;
2. A traffic composition part describing propagation of traffic composition for substreams with different destinations; and
3. A dynamic assignment part, which routes traffic substreams so as to guarantee dynamic user optimum conditions in real time.

The dynamic assignment part may be used both for modeling and for control purposes (e.g., in the context of a route guidance system). A feedback concept is proposed for development of the dynamic assignment part (Item 3).

DYNAMIC MODELING OF TRAFFIC NETWORKS

A precise mathematical framework for deterministic, macroscopic modeling of traffic networks has been published elsewhere (*2*). Therefore, in this paper only the basic approach and the resulting model structure will be outlined.

Definitions

Consider a traffic network represented by a directed graph. (See Figure 1.) N and M denote the sets of network nodes and links, respectively. Let I_n and O_n be the sets of links entering and leaving the nth node, respectively. It is assumed that traffic demands d_{ij} (veh/h) arriving at origin nodes $i \in D$ and being routed through the traffic network to some destination nodes $j \in S$, where D and S denote the sets of origin and destination nodes, respectively. A node may belong either to D or to S or to both or to none of them. The traffic demand being routed through the network is denoted by q_m and Q_m, $m \in M$, the traffic volume (veh/h) entering and leaving the link m, respectively. In each link there may be traffic subflows destinated to different destinations j. We denote by γ_{mj} (Γ_{mj}), $j \in S_m$, the composition rate, that portion of q_m (Q_m) destinated to Node j, where S_m is the set of destination nodes, which are reachable via Link m. Note that

$$\sum_{j \in S_m} \gamma_{mj} = 1, \sum_{j \in S_m} \Gamma_{mj} = 1 \qquad (1)$$

always holds. Hence, the number of independent composition rates for a link is equal to the cardinality of S_m minus one.

Modeling of Network Nodes

A model of a network node n should be capable of calculating q_m, γ_{mj} for $m \in O_n$ on the basis of d_{nj}, $j \in S^n$, and of Q_m, Γ_{mj}, $m \in I_n$, where S^n is the union of all S_m, $m \in O_n$. To make this possible, an additional variable that reflects the route choice behavior of drivers needs to be introduced. Hence, the splitting rates β_{nj}^m, $m \in O_n$, denote the portion of traffic flow which arrives at node n (regardless its origin), is destined to j, and is exiting node n by link m. In other words, splitting rates are turning rates by destination. Note that

Lehrstuhl für Steuerungs- und Regelungstechnik, Technische Universität München, Postfach 20 24 20, D-8000 München 2, FR Germany.

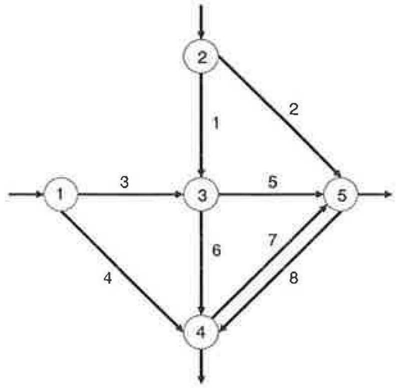


FIGURE 1 Example network.

$$\sum_{m \in \Lambda_{nj}} \beta_{nj}^{m} = 1,\ 0 \leq \beta_{nj}^{m} \leq 1 \tag{2}$$

always hold, where Λ_{nj} is the set of output links of node n for which $j \in S_m$ holds. Hence, if λ_{nj} is the cardinality of the set Λ_{nj}, the number of independent splitting rates is $\lambda_{nj} - 1$ for each couple (n, j).

With this definition we obtain at each node $n \in N$

$$q_m = \sum_{j \in S^n} \beta_{nj}^{m} q_{nj} \qquad m \in O_n \tag{3}$$

$$\gamma_{mj} = \beta_{nj}^{m} q_{nj}/q_m \qquad m \in O_n, j \in S^n \tag{4}$$

where q_{nj} is the traffic volume (veh/h) arriving at node n (regardless of its origin) and destinated to j, that is

$$q_{nj} = \sum_{m \in I_n} Q_m \Gamma_{mj} + d_{nj} \qquad j \in S^n \tag{5}$$

Note that Equations 3 through 5 distribute the traffic flow entering a network node among the leaving links according to the destination of the involved subflows and according to the splitting rates β_{nj}^{m} (see Figure 2).

Modeling of Network Links

The evolution of the traffic state inside a link and at the link's output depends entirely upon the traffic conditions at the link's boundaries. If the variables q_m, γ_{mj}, $m \in M$, $j \in S_m$, are organized in an input vector $\underline{U}$ and the variables Q_m, Γ_m, $m \in M$, $j \in S_m$, in an output vector $\underline{Y}$, the very general link model structure can be written as

$$\underline{Y}(k) = \underline{G}[\underline{x}(k), \underline{U}(k)] \tag{6}$$

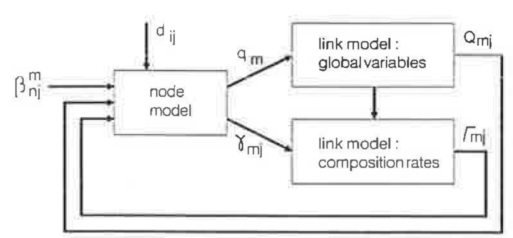


FIGURE 2 The overall network model.

$$\underline{x}(k + 1) = \underline{F}[\underline{x}(k), \underline{U}(k)] \tag{7}$$

where $k = 0, 1, 2, \ldots$ is the discrete time index, that is, $\underline{x}(k) = \underline{x}(k \cdot T)$, T being the sample time interval. The dimension and the composition of the state vector $\underline{x}$ depends upon the particular link model used.

For the development of dynamic link models, the global traffic variables that do not depend upon a particular destination are concentrated on first. A relatively simple dynamic model utilized by Merchant and Nemhauser (*3*) and by Wie (*4*) requires introduction of traffic density ρ_m (veh/km) in link m and makes use of the conservation equation

$$\rho_m(k + 1) = \rho_m(k) + (T/\Delta_m)[q_m(k) - Q_m(k)] \tag{8}$$

where Δ_m is the length of link m. Furthermore it is assumed that Q_m is given in terms of ρ_m by a nonlinear algebraic relation. As an example, consider the exponential relationship.

$$Q_m(k) = q_{\max,m}\,[1 - \exp(-\rho_m(k)/R_m)] \tag{9}$$

where $q_{\max,m}$ and R_m are constant parameters. For stability reasons, the sample time interval should be chosen such that $T < \min(\Delta_m R_m/q_{\max,m}, m \in M)$. Equations 8 and 9 provide the form required by Equations 6 and 7.

The choice of a dynamic link flow model depends upon the physical background of the corresponding network link. Equations 8 and 9 are very similar to the platoon dispersion model used in TRANSYT for links that represent urban streets [see Robertson (*5*)]. If some of the network links represent freeway axes, more sophisticated dynamic models are required [see Papageorgiou (*6*)]. Sophisticated models of freeways links consider subdivision of links into a number of segments and apply hydrodynamic equations to each of the segments. A general freeway network modeling computer program based on the presented network framework and on sophisticated dynamic modeling of traffic flow along the links is now available (*7*).

Note that the network modeling framework presented here allows for using different models for corresponding groups of links, which is an essential feature in modeling corridor or other mixed-traffic networks.

A traffic network may include control inputs such as urban traffic lights or freeway ramp metering. For the sake of simplicity, these control inputs have not been considered in the general Equations 6 and 7 because the description of route choice phenomena in the network will be concentrated on. A suitable extension of the general modeling structure to include traffic control measures is given elsewhere (*2*).

The dynamic modeling of composition rates $\gamma_{mj}(k)$ along a link will now be considered. There is no sound theoretical basis for development of a macroscopic model that propagates the composition rates $\gamma_{mj}(k)$ along a link. A possible approximation reads

$$\Gamma_{mj}(k + 1) = \alpha_m \gamma_{mj}(k) + (1 - \alpha_m)\Gamma_{mj}(k) \tag{10}$$

where α_m may be either constant or dependent upon the travel time along the link, that is

$$\alpha_m(k) = T/\tau_m(k) \tag{11}$$

The travel time is given by $\tau_m(k) = \Delta_m/v_m(k)$, and the mean speed reads $v_m(k) = Q_m(k)/\rho_m(k)$. Further composition rate models are presented and discussed in Papageorgiou (*2*).

Integrated Dynamic Network Model

The overall network model consists of the following interacting modules (see Figure 2):

- The node modeling Equations 3 through 5, and
- The chosen link models for traffic flow and for composition rates.

The overall model can be expressed by the general nonlinear, discrete-time vector equation:

$$\underline{x}(k + 1) = \underline{f}[\underline{x}(k), \underline{\beta}(k), D(k)],$$

$$k = 0, \ldots, K - 1 \tag{12}$$

where $\underline{x}$ is the state vector and $D = (d_{ij})$ is the origin-destination demand matrix. The vector $\underline{\beta}(k) \in R^p$ includes all independent splitting rates and obeys

$$\underline{0} \leq \underline{\beta}(k) \leq \underline{1} \tag{13}$$

Figure 3 illustrates Equation 12 from a system theoretic viewpoint; $\underline{\beta}(k)$'s are input variables and $D(k)$'s are disturbances. Note that Equation 12 can be resolved for $\underline{x}(k)$, $k = 0, \ldots, K$, if the trajectories $\underline{\beta}(k)$, $D(k)$, $k = 0, \ldots, K - 1$, and the initial condition $\underline{x}(0)$ are given.

For example, using the dynamic Equations 8 and 10 for a link model and replacing the static Equations 3, 4, and 5 (node model) by Equations 9 and 11, a state vector consisting of traffic densities and composition rates is obtained. In the case of the example network presented in Figure 1, the corresponding state vector reads

$$\underline{x} = [\rho_1 \ldots \rho_8 \gamma_1 \ldots \gamma_6]^T \tag{14}$$

where $\gamma_1, \ldots, \gamma_6$ are the independent composition rates for links $1, \ldots, 6$. The p independent splitting rates for this example are listed in Table 1 (here $p = 6$).

Physical Significance of Splitting Rates

The independent splitting rates $\underline{\beta}$ reflect the drivers' behavior with respect to alternative route choice. Clearly, the drivers'

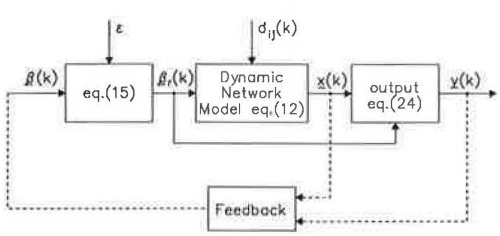


FIGURE 3 Network process considering compliance rate ε and feedback for dynamic assignment.

TABLE 1 SPLITTING RATES FOR EXAMPLE NETWORK

From Network Node	To Destination Node	Splitting Rates	Independent Splitting Rates	Number
2	4	β_{24}^1, β_{24}^2	β_{24}^1	1
2	5	β_{25}^1, β_{25}^2	β_{25}^2	2
1	5	β_{15}^3, β_{15}^4	β_{15}^3	3
1	4	β_{14}^3, β_{14}^4	β_{14}^4	4
3	5	β_{35}^5, β_{35}^6	β_{35}^5	5
3	4	β_{34}^5, β_{34}^6	β_{34}^6	6

behavior may be influenced by real-time information or route recommendation provided to them either by use of suitably located variable message signs or by individual communication with suitably equipped vehicles. The authors' interest in the independent splitting rates is twofold:

1. Modeling: How should $\underline{\beta}$ be calculated in absence of any communication to the drivers so as to reflect their natural behavior?
2. Control: If $\underline{\beta}$ is manipulable through suitable communications to the drivers, what is the best choice of $\underline{\beta}$?

The next sections present a feedback mechanism that leads to the specification of $\underline{\beta}$ so as to satisfy some generalized dynamic user optimal conditions.

Assume that $\underline{\beta}$ is manipulable by use of variable message signs which recommend route choice to the drivers. In this case, the modeling results of this section suggest that one variable message sign should be installed for each independent β_i. More specifically, at each node n of the network, the number of required variable message signs equals the number of destination nodes j, which are reachable from node n and for which a splitting at node n is possible.

The case where only a portion of the vehicles are equipped and/or only a portion of the drivers follow the recommendations provided will now be discussed. A parameter E, $0 \leq \varepsilon \leq 1$, reflecting the compliance rate and/or the rate of equipped vehicles such that for $\varepsilon = 0$ none follows the recommendations, and for $\varepsilon = 1$ everybody follows the recommendations. If β is the splitting rate ordered by the control system and β_r is the resulting real splitting rate, it may be written

$$\beta_r = 1 - (1 - \beta)\varepsilon \tag{15}$$

Equation 15 may be integrated into the general state space model (Equation 12). Figure 3 illustrates that—from a system theoretic viewpoint—ε can be interpreted as a disturbance acting on the process under control.

DYNAMIC USER OPTIMUM

Dynamic User Optimum Definition

For simplicity, the single-origin–single-destination network of Figure 4 will first be considered. A generalization of the obtained results will be considered later. The demand arriving at node 1 (Figure 4) is distributed between the links according

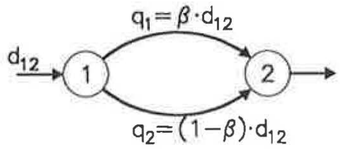


FIGURE 4 A single-origin-single-destination network.

to the independent splitting rate β such that $q_1 = \beta d_{12}$, $q_2 = (1 - \beta)d_{12}$.

As it is well known, steady-state user optimal assignment conditions for the single-origin–single-destination network of Figure 4 are present if and only if β obeys the following conditions

$$\begin{aligned} \beta &= 1 && \text{if } C_1 < C_2 \\ 0 < \beta &< 1 && \text{if } C_1 = C_2 \\ \beta &= 0 && \text{if } C_1 > C_2 \end{aligned} \tag{16}$$

where C_m is a measure of the individual cost along link m. For example, it can be assumed that C_m is the travel time along link m, in which case

$$C_m = \tau_m = \Delta_m/v_m \tag{17}$$

where v_m is the mean speed on link m.

Equation 16 may be readily expanded into a dynamic user assignment condition applying for $k = 0, \ldots, K$:

$$\begin{aligned} \beta(k) &= 1 && \text{if } C_1(k) < C_2(k) \\ 0 < \beta(k) &< 1 && \text{if } C_1(k) = C_2(k) \\ \beta(k) &= 0 && \text{if } C_1(k) > C_2(k) \end{aligned} \tag{18}$$

An equivalent form of these conditions may be expressed in terms of the quantity

$$y = \beta\Psi(C_1 - C_2) + (1 - \beta)\Psi(C_2 - C_1) \tag{19}$$

where the function Ψ is defined $\Psi(\cdot) = \max(0, \cdot)$. With this definition, the conditions (Equation 18) are equivalent to

$$y(k) = 0 \qquad k = 0, \ldots, K. \tag{20}$$

A precise mathematical description of dynamic user optimal conditions now requires an adequate definition of the individual cost $C_m(k)$ along link m at time k. For simplicity but without loss of generality, it will be assumed that the individual cost corresponds to a notion of travel time.

A particular dynamic generalization of the static user optimum is provided by the "reactive user optimum," which is defined by

$$C_m(k) = \tau_m(k) = \Delta_m/v_m(k) \tag{21}$$

By its definition, $\tau_m(k)$ depends upon the current traffic conditions on link m. In other words, $\tau_m(k)$ is an ideal travel time spent by an ideal vehicle that travels along the link m under traffic conditions that correspond to the current traffic conditions.

The reactive user optimum relies on the assumptions that

1. Traffic conditions in the network are not predictable because of, for example, incidents, variable demands, and stochastics.
2. Complete real-time information is available to the decision makers.

In fact, under these assumptions, a driver arriving at a bifurcation location, will choose the route which, according to reliable real-time information, currently appears to be shorter. A connection with an alternative, predictive dynamic assignment definition is discussed by Papageorgiou (*2*).

Generalization

The preceding statements will now be generalized for a multiple origin–multiple destination network, concentrating on the connection of a network node $n \in N$ with a destination node $j \in S^n$. There are three complications when compared with the simple network of Figure 4:

1. The number of output links of node n may be greater than two.
2. Each alternative route may consist of more than one link.
3. Some output links of node n may belong to more than one alternative route (because of farther downstream bifurcations).

As far as the first complication is concerned, it will be assumed that the number of output links of each node does not exceed two. This is without loss of generality because any node with more than two output links may be decomposed as indicated in Figure 5, by introducing artificial links without dynamics and with zero costs. This simplification gives $\lambda_{nj} \leq 2$ for all pairs ($n \in N$, $j \in S^n$). Eventually there is at most one independent splitting rate β_{nj} for each pair ($n \in N$, $j \in S^n$).

To handle the other two complications, the shortest travel time between nodes n and j through link $m \in \Lambda_{nj}$ is introduced.

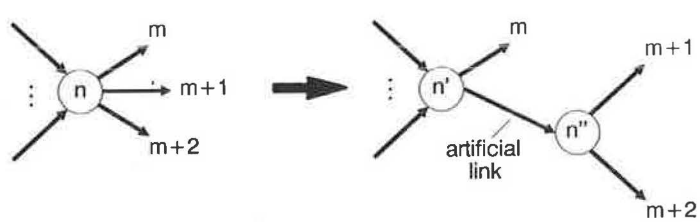


FIGURE 5 Decomposition of a complex node.

More precisely

$$\tau_{nj}^{m} = \min_{z \in Z} \sum_{\nu \in L_{nj}^{z}} \tau_{\nu} \qquad m \in \Lambda_{nj} \tag{22}$$

where $Z = \{z | m \in L_{nj}^{z}\}$.

Note that Equation 22 may be used as a generalization of Equation 21 for the general network case. Note further that the definition (Equation 22) implies the execution of a shortest path algorithm for the calculation of τ_{nj}^{m} from known link costs τ_{ν}.

As an example, consider the corresponding formulas for τ_{14}^{3}, τ_{14}^{4} in the example network of Figure 1:

$$\tau_{14}^{3} = \min \{\tau_3 + \tau_6, \tau_3 + \tau_5 + \tau_8\}$$

$$\tau_{14}^{4} = \tau_4$$

With these definitions, generalization of Equations 18, 19, and 20 is straightforward and the statements made for the simple network of Figure 4 apply to general networks as well. More precisely, perform the following replacements in Equations 18, 19, and 20:

β by β_{nj}, C_1 by τ_{nj}^{m}, C_2 by τ_{nj}^{1}, and y by y_{nj}

The resulting conditions are required to hold: $\forall\, k \in [0,K]$; $\forall\, n \in N$; $\forall\, j \in S^{n}$; $m,\mu \in \Lambda_{nj}$.

Note that there is exactly one variable y_{nj} assigned to each independent splitting rate β_{nj} of the general network. Thus a vector $\underline{y} \in R^{P}$ comprising all y_{nj}, $n \in N, j \in S^{n}$ may be defined. With the preceding generalizations it may be stated that

Reactive dynamic user optimal conditions in a general traffic network are present if and only if

$$\underline{y}(k) = \underline{0},\ k = 0, \ldots, K - 1. \tag{23}$$

Because the travel times $\tau_m(k)$ depend upon the system state $\underline{x}(k)$, general notation for the overall network (see Figure 3) may be written as

$$\underline{y}(k), = \underline{g}[\underline{x}(k), \underline{\beta}(k)] \tag{24}$$

There is no guarantee that there is a unique $\underline{\beta}(k)$ trajectory satisfying Equation 23 under a given demand in a given traffic network. Hence more than one solution may generally be present for the dynamic traffic assignment problem defined in this paper.

Dynamic User Optimum Via Feedback Regulation

The question to be treated in this section reads: is it possible to establish a dynamic user optimum by use of feedback regulation, that is, by a relationship $\underline{\beta}(k) = R[\underline{x}(k)]$? The significance of a real-time feedback law for dynamic assignment and for route guidance is obvious and will be further discussed later.

First note that Equation 18 is satisfied (and hence a reactive user optimum is reached) if the following simple feedback law is applied to a general traffic network.

$$\beta_{nj}(k) = \begin{cases} 1 \text{ if } \tau_{nj}^{m}(k) > \tau_{nj}^{\mu}(k) \\ 0 \text{ if } \tau_{nj}^{m}(k) < \tau_{nj}^{\mu}(k) \end{cases} \tag{25}$$

This feedback law is a bang-bang one, that is, the input variable $\beta_{nj}(k)$ takes values only on its bounds. Such a bang-bang controller may be adequate in the case of collective route guidance in which no values other than 0 and 1 can be implemented.

In some cases, a bang-bang solution may not be satisfactory. In fact, for individual route guidance with a high rate of equipped vehicles, bang-bang control may lead to strong perturbations of traffic flow. A smooth regulation may be achieved if $\underline{y}(k)$ is understood as the output of a process with input $\underline{\beta}(k)$ (see Figure 3). In this case, $\underline{y}(k)$ might be kept near or equal to zero by introducing a feedback law

$$\underline{\beta}(k) = \underline{\beta}(k-1) + K_p[\underline{x}(k) - \underline{x}(k-1)] + K_I\underline{y}(k) \tag{26}$$

where K_p is the proportional gain matrix and K_I is the integral gain matrix, which is assumed to have full rank p. Equation 26 describes a multivariable feedback regulator with integral parts.

To investigate the properties of the closed-loop system, it is first assumed (as a theoretical experiment) that the compliance rate ε demands to be constant, that is, $\varepsilon(k) = \bar{\varepsilon}$ and $D(k) = \overline{D}$, $k = 0, \ldots, K - 1$. If the closed-loop system is stable, a steady-state solution of Equation 26 then reads

$$K_I\underline{\bar{y}} = \underline{0} \tag{27}$$

where bars denote steady-state values. Since K_I is chosen to have full rank, $\underline{\bar{y}} = \underline{0}$ results from Equation 27 which is a well-known result in automatic control theory. Thus for constant demands and constant compliance rates, the multivariable feedback regulator (Equation 26) leads automatically to dynamic user optimum conditions without knowledge of the compliance rates and of the demands.

If disturbances $D(k)$, $\varepsilon(k)$ are not constant, as is usually the case, the feedback regulation will keep $\underline{y}(k)$ near zero for a reasonable choice of the gain matrices. It is important to underline again that the feedback laws (Equations 25 and 26) do not include any information on the present or future values of the demand and of the compliance rate.

The gain matrices K_p, K_I of the feedback law (Equation 26) should be chosen such that the overall closed-loop system be stable in a reasonable operating region around a theoretical steady-state. For example, specification of the gain matrices may be achieved by linearization of the system equations around a theoretical steady-state and by application of linear-quadratic (LQ) optimization methodology [see Papageorgiou (2) or other suitable methods; e.g., Kwakernaak and Sivan (8)]. Suitable gain matrices can be developed by the LQ method by means of a systematic trial-and-error procedure. Although not trivial, this development can be performed efficiently with some experience and basic knowledge of the LQ approach even for large-scale networks. The resulting LQ regulator is known to have excellent robustness properties for a wide range of process conditions.

The preceding results will now be illustrated on the basis of the example network of Figure 1 and the modeling Equations 8 through 11. Appropriate gain matrices K_p, K_I were selected by application of the LQ method, see Senninger (9) for details and see Papageorgiou (2) or Senninger (9) for the matrix values. The feedback law (Equation 26) was applied to the nonlinear network traffic modeling equations for different demand scenarios. First consider the rectangular demand scenario depicted in Figure 6. Figure 7 shows the resulting travel time differences $\Delta\tau_i(k)$ (in percent) for the six pairs of alternative routes corresponding to the six independent splitting rates of Table 1. Figure 8 depicts the corresponding trajectories of $\underline{\beta}(k)$. The following remarks may be stated:

1. The required condition $\underline{y}(k) = \underline{0}$ is satisfied for most $k \in [0,K]$ through the action of the feedback regulator. In fact, for most $k \in [0,K]$, either $\beta_i(k) = 1$ (no splitting) and $\Delta\tau_i(0) < 0$ (routes are not competitive), or $0 < \beta_i(k) < 1$ (splitting of the corresponding substream occurs) and $\Delta\tau_i(k) = 0$ (equal travel times on alternative routs) hold.

2. A steady-state is achieved for each of the three sets of constant demand values included in the demand scenario of Figure 6. Note that each steady-state is equivalent to a corresponding static user optimal equilibrium.

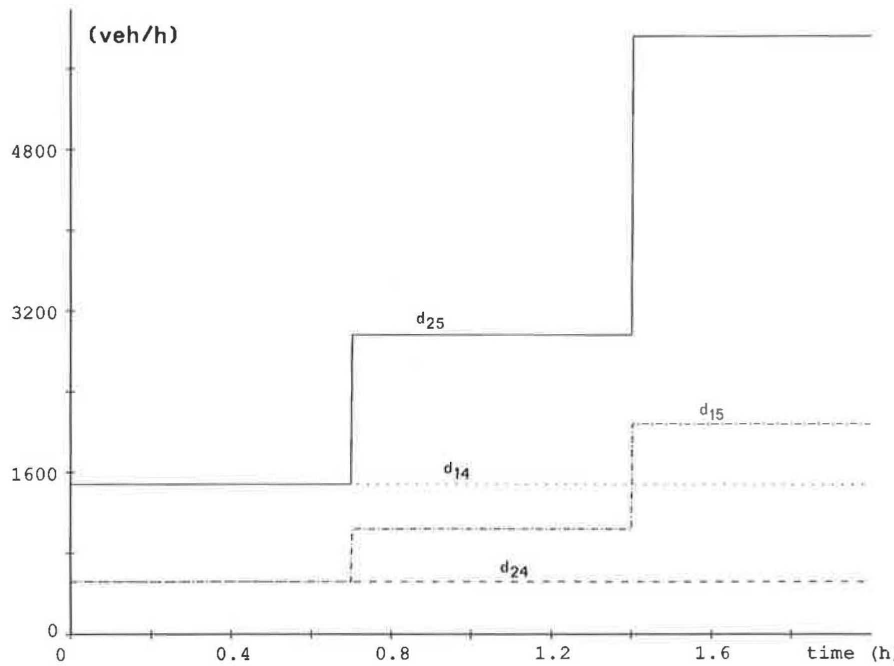


FIGURE 6 A rectangular demand scenario.

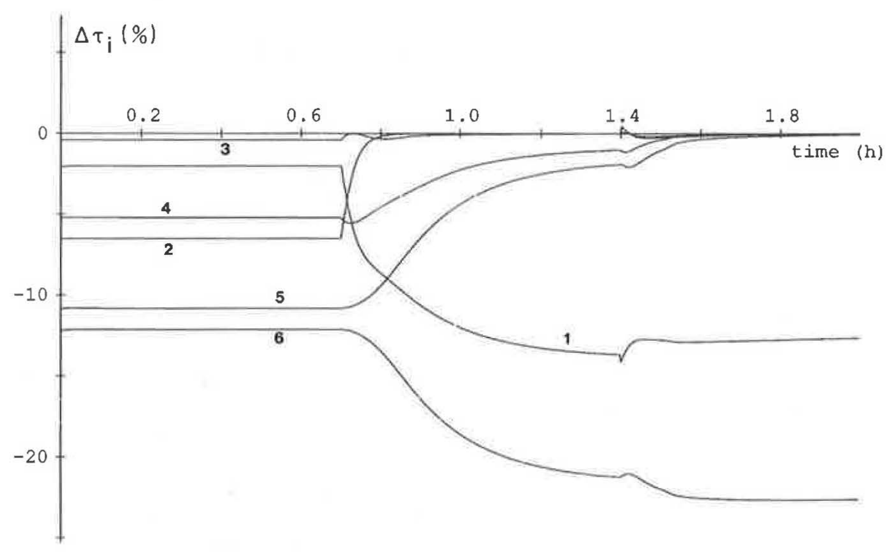


FIGURE 7 Travel time differences $\Delta\tau_i$ for six pairs of alternative routes for multivariable feedback law (Equation 26).

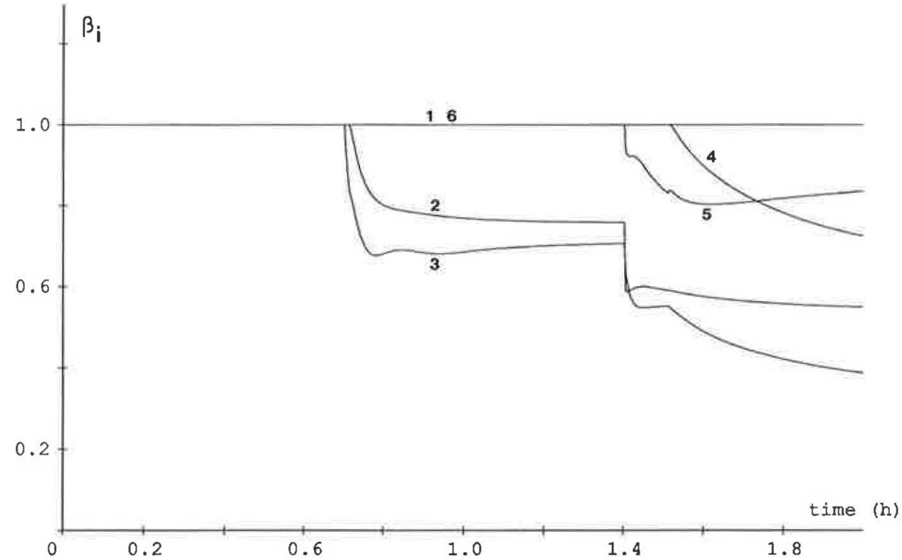


FIGURE 8 Splitting rates provided by multivariable feedback law (Equation 26).

Figure 9 depicts the travel time differences $\Delta\tau_i(k)$ resulting by application of the bang-bang controller (Equation 25) to the same network with the same rectangular demand of Figure 6. Note that the bang-bang controller

1. Leads to a slightly oscillatory behavior for some $\Delta\tau_i(k)$.
2. Equalizes travel times on alternative routes for a smaller number of alternative route pairs as compared to the multivariable regulator.

Figure 10 depicts a triangular demand scenario and Figures 11 and 12 depict the resulting travel time differences $\Delta\tau_i(k)$ and the splitting rates $\beta_i(k)$, $i = 1, \ldots, 6$, for the multivariable regulator. Again the feedback regulator (Equation 26) succeeds in keeping $\underline{y}(k)$ close (but not exactly equal) to zero although the demands $d_{ij}(k)$ are unknown to the feedback law. Figure 13 depicts the corresponding results of the bang-bang controller.

The advantages of the feedback concept when applied to traffic networks in the aim of establishing reactive dynamic user optimal conditions will now be summarized:

1. The feedback concept requires only few calculations at each time instant k. Moreover, it is a real-time procedure such that no iterations or other time consuming algorithms are required.
2. The feedback law does not utilize current or future values of the process disturbances of the origin-destination demands $D(k)$ and of the compliance rate ε (see Figure 3). Nevertheless, its sensitivity with respect to variations of these disturbances seems low as demonstrated by the preceding results.
3. Note that for a real-life control application, only the feedback portion of Figure 3 is implemented using measurements from the real traffic process, that is, no model calculations are required in real time.

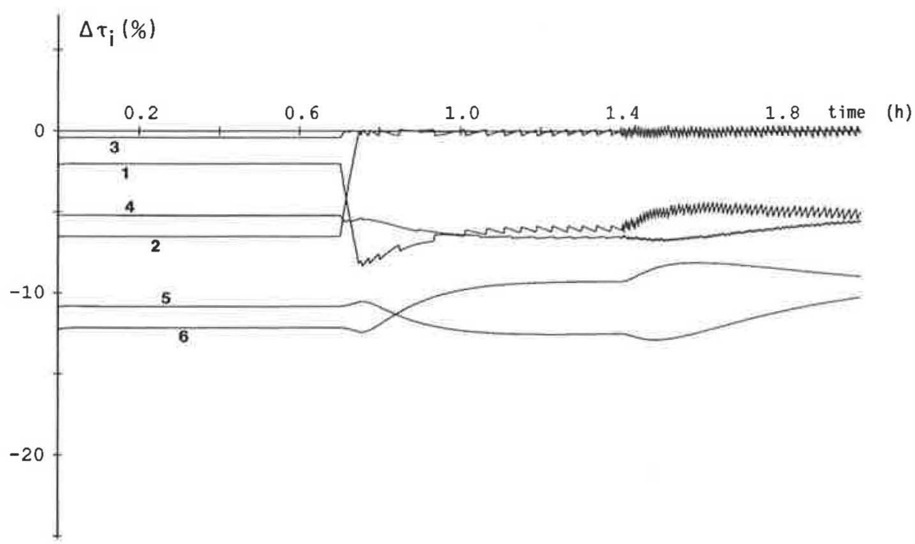


FIGURE 9 Travel time differences $\Delta\tau_i$ for bang-bang controller.

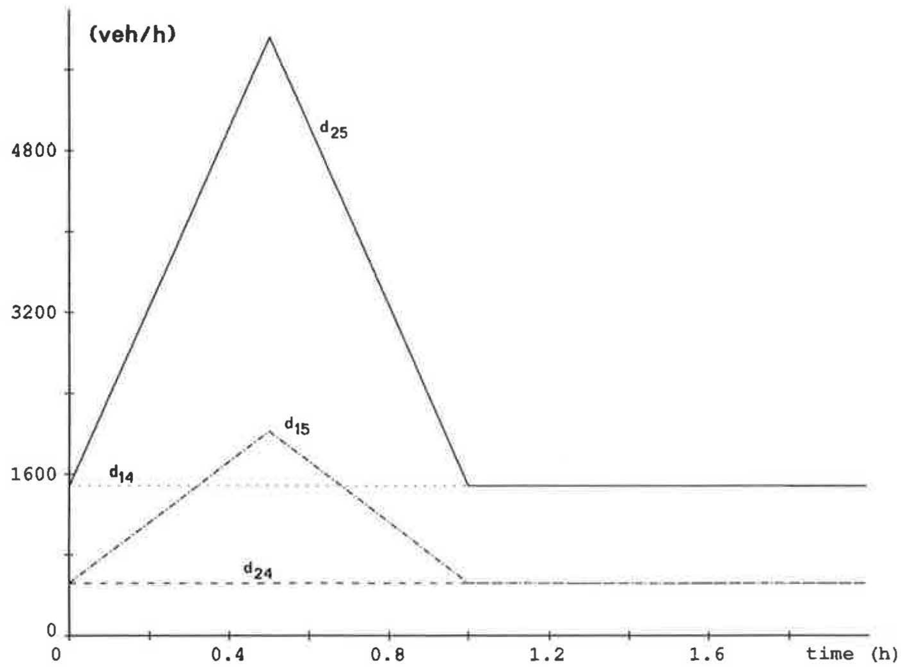


FIGURE 10 A triangular demand scenario.

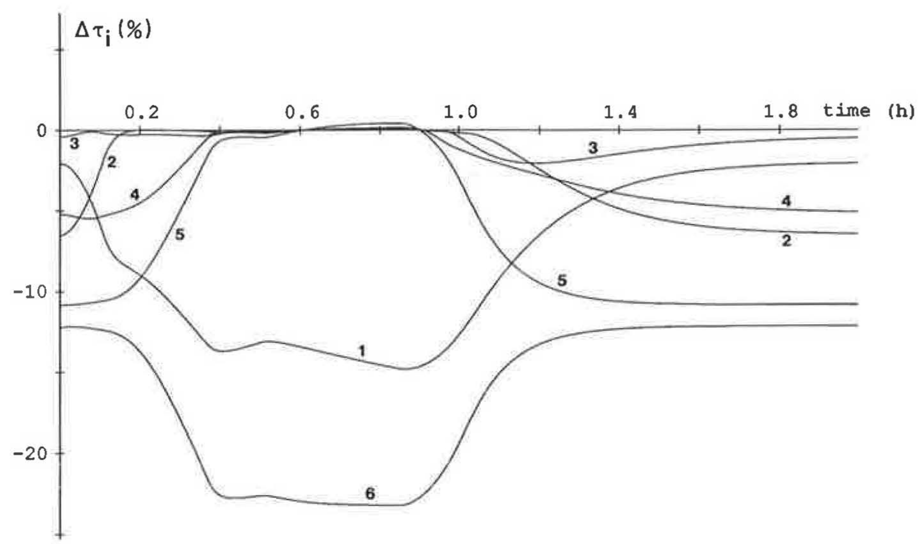


FIGURE 11 Travel time differences $\Delta\tau_i$ for triangular demand and multivariable feedback.

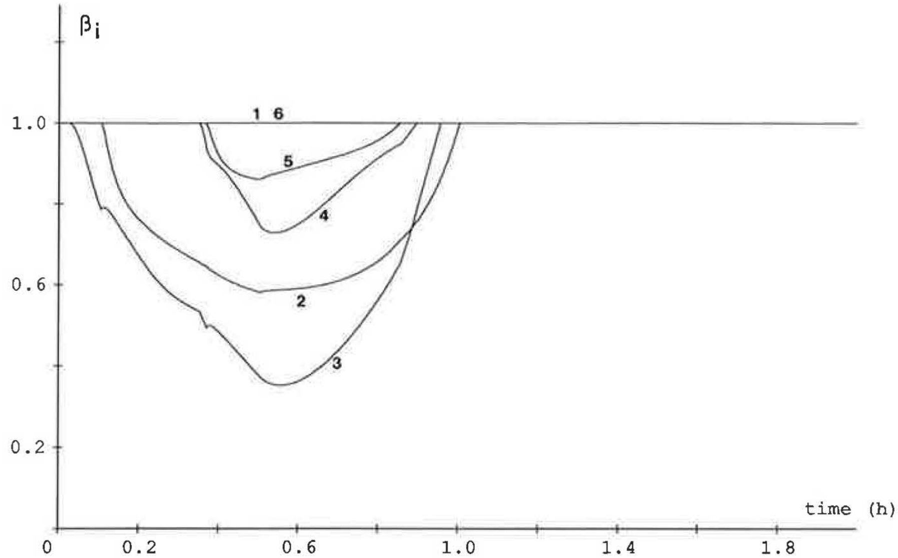


FIGURE 12 Splitting rates for triangular demand scenario and multivariable feedback.

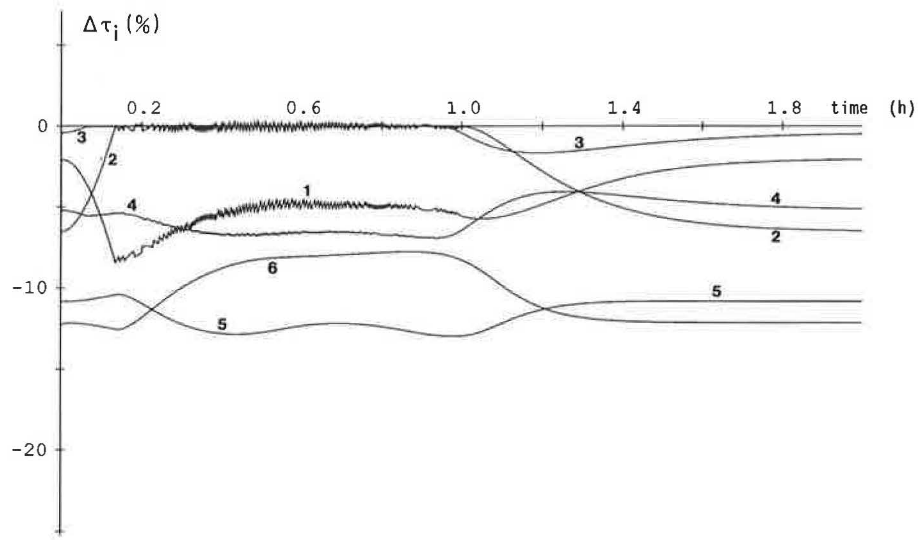


FIGURE 13 Travel time differences $\Delta\tau_i$ for triangular demand and bang-bang controller.

It should be emphasized that the proposed feedback concept is of a reactive character—it reacts indirectly to the disturbances (namely via their impact on the traffic state)—and this is the reason why it does not need disturbance predictions. For example, if the compliance rate ε is too low (high) this will have an impact on the traffic state and will lead automatically the feedback law to according modification of the input $\boldsymbol{\beta}$ so as to approach the goal $\underline{y}(k) \approx \underline{\mathbf{0}}$.

Under realistic conditions, with $D(k)$ and ε varying strongly with time, the feedback concept cannot lead exactly to $\underline{y}(k) = \underline{\mathbf{0}}$ but hopefully to $\underline{y}(k) \approx \underline{\mathbf{0}}$ as demonstrated in the example tests. Furthermore, for very strong variations of the traffic state from the linearization conditions (severe congestion!), the linear regulator may need a long time to lead the output $\underline{y}(k)$ near zero, although it will react in the right sense. But what could be an alternative approach? One should be able to predict the origin-destination demands and compliance rates (which is rather unrealistic) and to apply a mathematical traffic model (accuracy?) in order to calculate iteratively (effort!) the route recommendations so as to achieve $\underline{y}(k) = \underline{\mathbf{0}}$ in the computer (in real life?). In contrast to such an approach, the feedback philosophy is to react to real-life measurements rather than to rely on predictions and mathematical models.

The tests of this paper are certainly not significant for practical applications with more realistic models or under real-life conditions. Nevertheless they do provide a very encouraging first step towards application of the innovative concept of feedback to a fairly complicated traffic problem, which opens the way to consideration of more realistic conditions. Anyhow, the particular link model used for the reported tests is not simpler than the ones used in previous research work on traffic assignment as cited in the references. Investigations of the feedback concept for route recommendation by variable message signs on freeway networks is currently under way, see Wolf (*10*) using realistic high-order link models of freeway traffic like METANET (*7*).

CONCLUSIONS

A general framework for deterministic dynamic modeling and control of traffic networks has been presented under non-elastic but time-varying demand conditions. The traffic network may include both freeways and urban roads. The presented methodology may be readily extended to consider control measures like ramp metering and signal settings, see Papageorgiou (*2*).

A feedback concept has been applied to the traffic network to achieve dynamic user optimum conditions. Because of three fundamental features—low computational effort, low sensitivity with respect to unknown origin-destination demands and unknown compliance rates, and integrated design procedure—the feedback concept appears particularly attractive for a broad class of traffic control problems, which include [see Papageorgiou (*2*), for more details]:

- Dynamic network traffic modeling including traffic assignment;
- Static user optimal traffic assignment;
- Integrated strategy development for route guidance and traffic control systems;
- Development of optimal traffic control strategies subject to dynamic assignment conditions; and
- Development of feedback strategies for a variety of traffic control problems including route guidance.

ACKNOWLEDGMENT

This work has been partially supported by the DRIVE Program of the European Commission under the Project No. V 1035 (CHRISTIANE). The authors would like to thank Hermann Senninger for essential contributions to the computational results reported in the paper.

REFERENCES

1. G. C. D'Ans, and D. C. Gazis. Optimal Control of Oversaturated Store-and-Forward Transportation Networks. *Transportation Science*, Vol. 10, 1976, pp. 1–19.
2. M. Papageorgiou. Dynamic Modeling, Assignment, and Route Guidance in Traffic Networks. *Transportation Research*, Vol. 24B, 1990, pp. 471–495.
3. 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.
4. B. W. Wie. An Application of Optimal Control Theory to Dynamic User Equilibrium Traffic Assignment. Presented at the 67th Annual Meeting of Transportation Research Board, Washington, D.C., 1988.
5. D. I. Robertson. "TRANSYT" Method for Area Traffic Control. *Traffic Engineering & Control*, Vol. 2, 1969, pp. 276–281.
6. M. Papageorgiou. *Application of Automatic Control Concepts to Traffic Flow Modeling and Control.* Springer-Verlag, New York, 1983.
7. A. Messmer and M. Papageorgiou. METANET: A Macroscopic Simulation Program for Motorway Networks. *Traffic Engineering & Control*, Vol. 31, 1990, pp. 466–470.
8. H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems.* New York, John Wiley & Sons Inc., 1972.
9. H. Senninger. *Regelung der Fließvorgänge in Verkehrsnetzen durch Wechselwegweiser.* Diploma Thesis No. 826 at Lehrstuhl für Steuerungs- und Regelungstechnik, Technical University of Munich, 1989.
10. G. Wolf. *Simulationstechnische Untersuchung von regelungstechni- schen Konzepten zur Wechselwegweisung in Schnellstrassennetzen.* Diploma Thesis at Lehrstuhl für Steuerungs- und Regelungstechnik, Technical University of Munich, 1990.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

Using an Interactive Route-Choice Simulator to Investigate Drivers' Compliance with Route Guidance Advice

PETER BONSALL AND TIM PARRY

Possible sources of data on drivers' reactions to route guidance are discussed. Field evidence is sketchy and appears likely to remain so for some time. It is argued that an interactive route-choice simulator might provide acceptable substitute data. The design and development of such a simulator, interactive guidance on routes (IGOR), is described; users make a series of journeys through test networks by indicating their desired exit from each junction they reach. At each junction IGOR displays a plan giving information about road sizes and alignments, signposts, current traffic conditions, and so on. For some journeys the user has access to a map of the network, guidance advice, or both. The advice system replicates in-vehicle systems, which advise the driver what exit to take at each junction in order to minimize journey time in the current traffic conditions. To ascertain the effect of variations in the quality of guidance on user response, a "wrong" exit is sometimes recommended. The use of IGOR to collect data under the Dedicated Road Infrastructure for Vehicle Safety in Europe initiative is described and important results are presented. It is observed that acceptance of an item of advice depends on its (objective) quality, the quality of previously received advice, the drivers' knowledge of the network, and on the extent to which advice is corroborated by other evidence. Compliance with advice is a function of its credibility and this in turn depends on past experience, local conditions, and psychological factors. The value of IGOR and its results are discussed. Plans for use and further analysis of the IGOR data are outlined together with some options for further development of the concept.

In-vehicle route guidance (IVRG) or information systems are under development in various parts of the world and some have already been implemented, albeit on a limited scale. Examples include directional aids (Etak's Navigator); real-time traffic information transmissions (via car radios using the HAR ARI or RDS systems); real-time congestion displays (General Logistics' Trafficmaster); guidance based on historic data (Mercedes-Benz's, Routen-Rechner); and guidance based on real-time data (Siemens' Ali-Scout).

It is widely believed that such systems will be popular with car drivers and that they will influence route choice. Theoretical calculations have suggested that the net effect could be to increase network efficiency significantly by improving the efficiency of individual drivers' routes and, perhaps, by deliberately seeking to modify individual route choices in the interests of an overall network optimum. These calculations assume that equipped drivers will follow the guidance given.

Attitudinal surveys in the United Kingdom, France, and Germany (*1*) have, however, suggested that, particularly when driving on familiar routes, many drivers might be reluctant to accept guidance from a computer even if it did purport to "know" about current traffic conditions. There is clearly a potential credibility problem because many drivers believe that their own knowledge of the network would be superior to that of a computerized guidance system. To some extent, of course, the drivers may be right—for example, even if no deliberate attempt is being made to sacrifice individual benefits in the interests of a network optimum, the system may simply not "know" about potential short cuts through back streets and may be basing its guidance on information about traffic conditions received some minutes in arrears of what is actually happening on the streets.

An understanding of how drivers are likely in practice to react to route guidance or information is clearly crucial. Without it any estimates of the impact of such systems on network performance will be of purely theoretical interest and detailed work on the design aspects, such as multirouting guidance algorithms to avoid feedback problems, will be flawed. The problem, however, is that data on driver response to guidance and information systems is not readily available.

POSSIBLE SOURCES OF DATA ON DRIVER RESPONSE

A data set that would allow the exploration of the influence of the following would be ideal: (*a*) system variables such as quality of guidance received and completeness of the guidance network; (*b*) situational circumstances including current traffic conditions; and (*c*) driver characteristics such as age, sex, driving experience, and familiarity with the network. Ideally, again, data would be in sufficient volume to enable the calibration of quantitative models.

Direct observation of the behavior of drivers equipped with in-vehicle guidance, such as that used in Berlin's LISB project (an implementation of Ali-Scout), would seem to be an attractive potential source of data. It is theoretically possible, under LISB, to log the routes chosen by equipped drivers and to compare these with the routes used by those drivers before receiving guidance and with the routes which they were advised to use. With the cooperation of the SNV Consultancy, and with full permission of the drivers themselves, preliminary analyses of the LISB records were conducted by the University of Leeds during 1989. Unfortunately, however, the centrally held data on individual vehicles' routes proved to be unreliable and the "automatic" monitoring of individual driv-

Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, U.K.

ers' response to guidance was found not to be a practical proposition. Another attempt to undertake automatic monitoring will be made in conjunction with the trial of the PATHFINDER route guidance system under California's PATH initiative.

Although automatic monitoring had proved to be impractical, the LISB trial was still an important potential source of data on driver response; in parallel with a much larger evaluation program being undertaken by the SNV consultancy, the University of Leeds was able to conduct a series of questionnaire surveys among LISB users (*2,3*). The questionnaires sought aggregate information on driver behavior (e.g., what proportion of drivers claimed to be seeking or following guidance in specified circumstances? What proportion have changed their routes as a result of guidance?) and attitudes (e.g., what reasons are quoted for not following advice?). The surveys yielded some valuable insights. Note for example that acceptance of advice was much lower on familiar journeys and that it declined over time. Among the most frequently quoted reasons for not following advice were that they thought it was sending them in the "wrong" direction and that they saw no reason to accept its advice to leave a normally good route, which had no obvious problems on the day in question. More complete results are quoted elsewhere (*4,5*).

Interesting and illuminating though these results from the LISB questionnaires may be, they do not provide the detailed data required for modeling purposes on issues such as the influence of the network specification, or the "accuracy" of guidance on driver response. This is primarily because they are subjective and aggregate rather than objective and disaggregate, but also because the LISB trial was designed primarily to prove the technology rather than to provide a test bed to explore such issues (*6*). Because there was no immediate prospect for such experimental designs being implemented in the field, or of disaggregate data being derived from them, it was necessary to consider the alternative methods of obtaining the needed data.

Because of the Institute's access to cars equipped with Ali-Scout, consideration was given to inviting "guinea-pig" drivers to exchange their own cars for ones equipped with Ali-Scout and then observing in what circumstances they did, or did not, follow the guidance received. Once again, however, it was concluded that it would not be possible by this means to explore all the variables of interest (particularly exposure to advice of differing qualities). It was also clear that cost and time considerations would limit the research to a vanishingly small sample of drivers.

Also considered was the use of relatively free-format interviews seeking attitudinal information from drivers who had been briefed, perhaps through a video, on the concept of route guidance. It was felt, however, that the resulting data would not have the precision required and that, because the issues had already been discussed in interviews in previous phases of our research, little new of substance would be learned.

The desiderata pointed to some form of stated-preference experiment wherein respondents would be offered a series of hypothetical route-choice decisions (with each option defined in terms of variables such as type of road, alignment relative to the destination, degree of congestion, whether it was the signposted route, and whether it was the advised route) and asked to indicate which option they would select. Such a technique could be administered using a conventional questionnaire or on an interactive computer (the advantages of which are discussed elsewhere) (*7,8*). On reflection however, no satisfactory way could be seen to build the quality of advice, or the drivers' previous experience of advice, into such an experiment. (Simply to tell respondents that the advice was right n times out of 100 would not suffice because it begs the question of how "right" is defined and how, in reality, a driver would perceive it.) It was this problem with conventional stated preference (SP) methods that led to developing the interactive simulator approach.

The value of interactive simulation as a means of gathering data on traveler response is increasingly recognized. Previous examples have included the Oxford work with HATS to study activity scheduling (*9*) and the Leeds use of multistage questionnaires to calibrate microsimulation models of car sharing schemes (*10*). Current interest has been stimulated by Mahmassani's work in Austin, Texas, on departure time choices and route-and-departure-time joint choices (*11–13*).

IGOR (interactive guidance on routes), a new model, provides drivers with feedback on the consequences of their own decisions but does not consider supply side response nor the consequences on other drivers—it is purely a device for gathering data on drivers' responses to the situations met.

INTERACTIVE SIMULATOR

Description of IGOR

IGOR runs on an IBM or compatible PC. Each user is invited to make a number of journeys through hypothetical networks by progressing from one junction to the next. At each junction, the participant is shown a plan of the junction, annotated with contextual information (see Figure 1 for a typical screen), and is invited to press a key to indicate the chosen direction. The participant is, on some of the journeys, provided with route guidance advice (in the form of a flashing arrow on the advised direction) but is free to ignore it if desired.

Each user makes several journeys from specified origins to specified destinations. For some journeys, a hard-copy map of the network is provided, for others, no map is provided. The conditions faced and the decision made at each junction are logged for subsequent analysis—thereby enabling the determination of the circumstances in which guidance is accepted or rejected. To examine the effect of the quality of advice on its acceptability, IGOR is programmed to provide a given amount of "bad" advice to participants. Some participants get better advice than others (the quality of advice received by each participant at each junction is known).

The current version of IGOR was developed over a 6-month period from late 1989 and has a number of features that should perhaps be described in greater detail.

Network

A hypothetical network containing 30 two-way links and 19 nodes has been the basis of the work. (See Figure 2 for a copy of the network map provided to users.) This network represents a typical small town with a historic center and a bypass

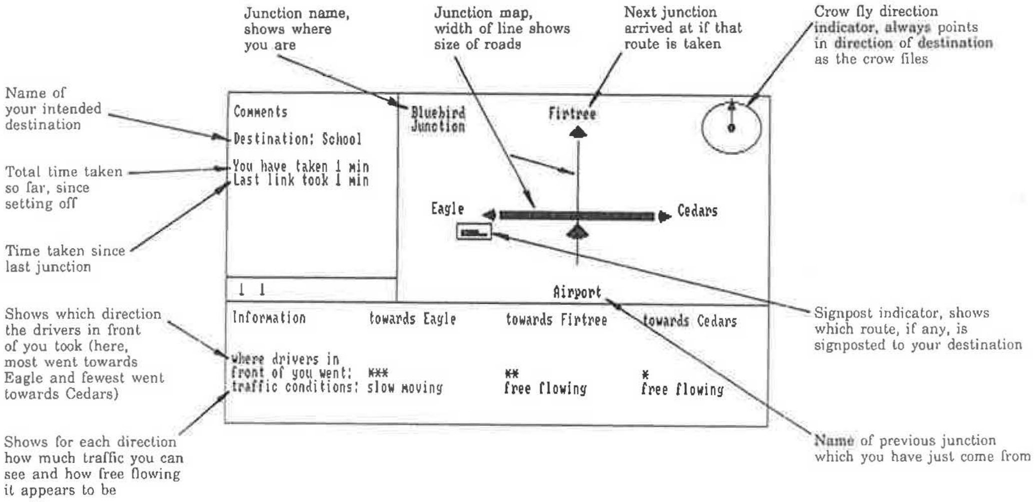


FIGURE 1 Annotated copy of IGOR's on-screen display.

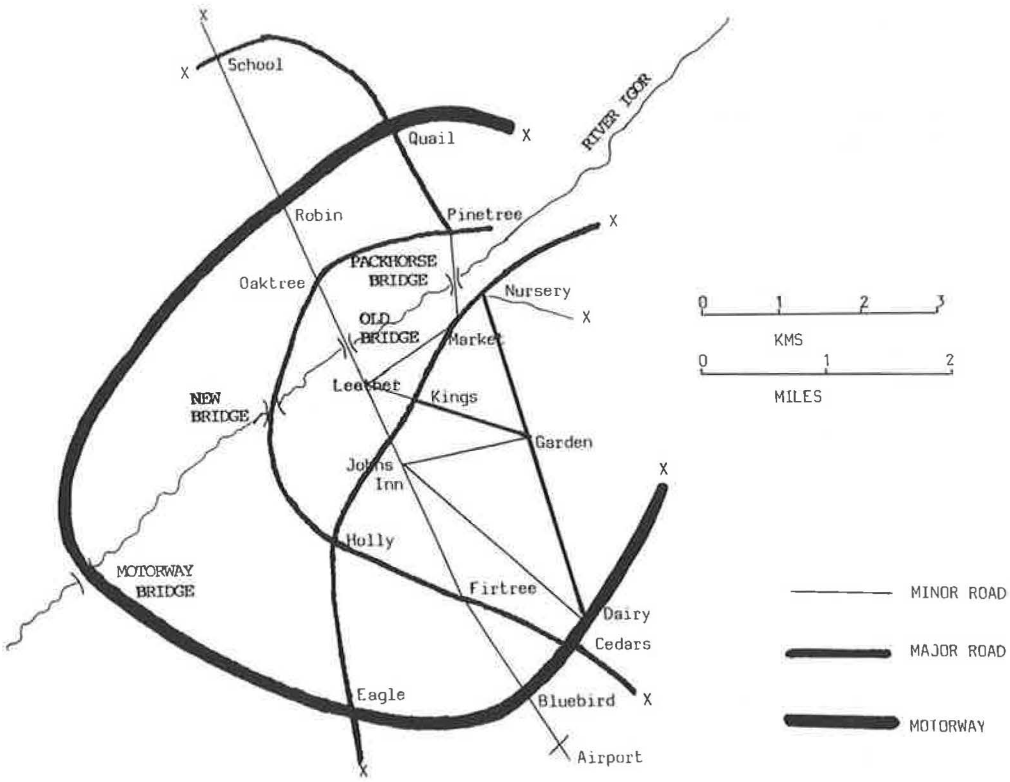


FIGURE 2 Map of network provided to IGOR users.

and has been designed to offer a number of interesting route-choice options. The junctions have been named thematically to introduce some sense of route identity.

Each link has a basic length but its traverse time during a particular journey depends on the supposed time of day and weather conditions and on a random element to represent day-on-day variability; thus the level of congestion varies from one journey to another to a degree that is realistic without being predictable.

IGOR can accept different network data and could, for example, represent one way streets or limited access junctions. There is no reason in principle why the network description should not be tailored to represent a real network with which the participant would be familiar. This option will be experimented with in due course.

Guidance System

The guidance system is currently programmed to produce the minimum time route to the specified destination given the current conditions. However, as is the case with real guidance systems such as Autoguide or LISB, IGOR's guidance system is "unaware" of some of the links in the network. This fact, together with the deliberate provision on some occasions of "bad" advice, ensures that the advised route is not always the quickest. This fact was not drawn to the attention of users.

On-Screen Display

Figure 1 shows a typical display during a journey. It reminds the participants of their destination, how long it has taken so far, and how long the last link took. The junction plan is aligned such that the entry arm is at the bottom. All the exits are labeled, to show which junction they lead to. The road type (highway, main street, or side street) is indicated by the width of the exit arm. The general crow-fly direction to the destination is indicated by an arrow (this is a proxy for the driver's general sense of direction). If it is desirable to indicate that the destination is signposted from the current junction, then the signpost symbol appears next to the appropriate exit arm. If guidance advice is being given, then the appropriate exit arm arrow will flash.

Traffic information is summarized in a table under the junction plan. It indicates what might be seen from the junction—thus, it indicates how much, if any, congestion can be seen on each exit link, and which exits are being chosen by drivers supposedly in front of the participant. (This information is included because previous research had suggested that drivers might be influenced, in their choice of exit arm, by what other drivers seemed to be doing.)

Other information that is displayed at specific junctions includes progress confirmation information such as "you have just entered the city center" or "you have just crossed the river." To provide some feedback, the driver is given, at the beginning of each journey, an estimate of the probable journey time assuming average travel conditions at that time of day and, at the end of each journey, the driver reminded of this estimate and told how long the trip actually took.

Sound Generation

In the current version of IGOR an engine sound is emitted as the driver moves from one junction to the next. The duration of the sound is proportional to the time required to traverse the link and its pitch is proportional to the speed. Thus the driver gets an impression of the passage of time and of the travel conditions—a high gear sound would accompany a trip round in uncongested ring road while a series of short low gear sounds would accompany a trip through the congested city center.

Interactive Questionnaires

As background for analysis of participants' decisions, information is required about their personal characteristics and attitudes. When participants first log on, they are asked to provide a certain amount of personal information (age, sex, home location, car ownership, access to company car, distance driven per year, whether they drive to work, whether they drive in the course of work, and how adept they consider themselves to be at finding new destinations for the first time). Also, before each journey, IGOR asks whether users already have an idea of which route they intend to take and if so, which bridge they intend to cross—this question is carefully phrased to minimize the possibility that they might subsequently feel committed to that route but does, it is thought, invite users to do a certain amount of strategic planning such as is often done in practice. The intention is to experiment with the inclusion and exclusion of this question in due course.

After completing their last journey, participants are presented with 6 SP questions in which they have to indicate which of two directions they would take in each of the 6 specified situations. The situations are designed such that in the first one the guidance system is in conflict with all other evidence, in the second it is in conflict with everything except the compass direction and so on, for each of the 5 variables. After these SP questions have been answered, players are asked some attitudinal questions: had they previously heard of in-car route guidance? (If so) had they expected to be useful to them? Had IGOR caused them to change their opinion for the better or worse? In what circumstances, or combinations or circumstances, would they, in real life, reject guidance? What criteria do they usually use in selecting routes in various situations?

Storage of Results for Subsequent Analysis

In designing IGOR facilitation of the analysis of the data was sought. Data for each participant is stored in a file with a unique identifier to record the time and data of the session. Each file contains a record for each decision made by the participant. Each record contains the participant's personal characteristics and answers to the attitudinal questions along with the description of the situation faced at the junction, information about the quality of advice received at this, and previous, junctions, and the decision actually made. The data is thus ready for analysis without any need for extensive file editing. The only data that needs to be brought in from a separate source are those relating to the conditions under which IGOR was used (i.e., whether any survey staff were present—and if so, who? And how the participant was "recruited"?). There is no reason in principle why this data too should not be typed into the PC and automatically entered into the files.

This description of the IGOR model summarizes one contained in a previous paper (*14*), which also includes more details of potential further developments of the concept.

Organization of an IGOR Session

There are, of course, many ways in which an IGOR session might be organized. The program will run on a portable PC and so can be used in peoples' homes or workplaces or at airports or transit stations much as one would a conventional questionnaire.

A typical session will have the following components:

1. Introduction—explanation of how to use IGOR;
2. Characteristics—questions on personal characteristics;
3. Familiarization—3 journeys without guidance designed

to familiarize the user with IGOR and the network (a map of the network is provided);

4. Description of guidance—introduction to the concept of in-car guidance; explanation that the guidance is based on current traffic conditions;

5. Reaction to guidance—6 journeys with guidance (using the network introduced in phase 3) and 3 journeys with guidance in a different network (no map provided), these represent unfamiliar journeys;

6. SP exercise—6 stated preference questions; and

7. Attitudes—direct questions on the perceived usefulness of guidance and on the user's normal route choice criteria.

All seven phases can be carried out through the PC screen and keyboard and could, in theory, be conducted without any survey staff in attendance. Indeed it would be possible to send out a disk containing the program to people with access to a PC for them to use it at their own convenience and then mail back the disk containing the data.

RESULTS OF AN ANGLO-FRENCH IMPLEMENTATION OF IGOR

Data Collection and Analysis

As part of a project under the European DRIVE initiative, IGOR has been used on behalf of the CARGOES consortium to collect information on drivers' reactions to route-guidance advice. A French translation of the on-screen information enabled French partners, INRETS, to use IGOR in Paris while it was being used in various locations in the United Kingdom. Some 350 participants were recruited, mainly through their employers, and most sessions took place at the participants' workplaces during early summer 1990. An analysis of participants' characteristics shows them to have been fairly representative of the car driving population.

Each participant made several journeys and each journey consisted of several decisions. The resulting data base contains data on more than 11,000 decisions. Further details of the data and its analysis can be found in the project report (*4*).

The IGOR data base has proved to be a very rich source of information, the analysis of which is by no means complete. It is, however, appropriate to present here those results relating to the subject of the current paper—drivers' acceptance or rejection of route-guidance advice.

Results

General

Analysis of the IGOR data base shows that, overall, about 70 percent of advice was accepted. The current analysis seeks to determine the extent to which acceptance or rejection is a function of objectively defined characteristics of the advice or of the decision makers.

Acceptance of Advice as a Function of Quality

As has been mentioned previously, unknown to the participants, the quality of advice given by IGOR was deliberately varied. The relationship between acceptance of an item of advice and its quality (defined as the minimum time to reach the destination by means of the advised route divided by the minimum time to reach it by any route) was examined. Plots of acceptance versus quality as shown in Figure 3. The x-axis in Figure 3(a) is an index of quality based on travel times in the IGOR network as they were at the time the journey was actually made, whereas the x-axis in Figure 3(b) is an index of quality based on free-flow travel conditions.

Both plots show that acceptance declines as the quality of advice decreases. It is clear that, although they were not in-

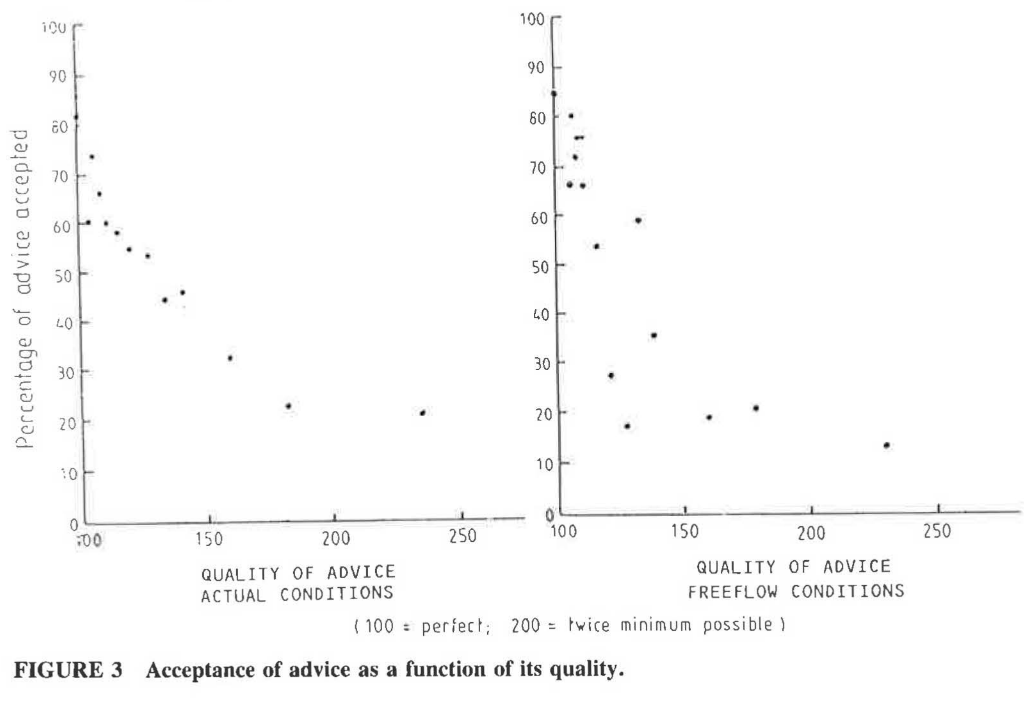


FIGURE 3 Acceptance of advice as a function of its quality.

formed that the quality of advice was variable, participants appear to have detected the fact and acted accordingly. The decline in acceptance is particularly sharp as the index of quality falls from about 100 (perfect advice) to about 150 (advised route half as long again as the best possible route). This initial decline is particularly strong in Figure 3(b), suggesting perhaps that participants' perception of the usefulness of advice was strongly conditioned by the physical layout of the network. This question will be revisited later in the paper.

Regression curves were fitted to the data shown in Figure 3 and the resulting equations were

$$p = 1.89 - 0.01q + 3.14qa^2 \qquad (r^2 = 0.94) \qquad (1)$$

$$p = 6.37 - 0.95qf + 0.00qf^2 + 8.2qf^3 \qquad (r^2 = 0.80) \qquad (2)$$

where

p = probability of acceptance,
qa = index of quality based on actual travel times,
qf = index of quality based on free-flow travel times, and
r^2 = squared correlation coefficient (fit of curve with data).

The poorer fit for the equation based on free-flow travel times may reflect the fact that some journeys were made in networks for which no map was available and of which the participant could not therefore be expected to have a good image of physical layout. This hypothesis will be pursued in further tests.

A third definition of quality, based not on the ratio of times via the advised and the true best route, but on the absolute difference between their times, was calculated and acceptance of advice was plotted against it. A relationship was apparent but the fit was not good.

Acceptance as a Function of Quality of Previous Advice

Acceptance of less than perfect advice by an individual seemed to depend not only on the quality of advice in question but also on the quality of advice previously received by that individual. Figure 4 shows that if previous advice had, on average, been very good, then even a very poor piece of advice was likely to be accepted, but if the quality of previous advice had, on average, been bad then a very poor item of advice was almost certain to be rejected.

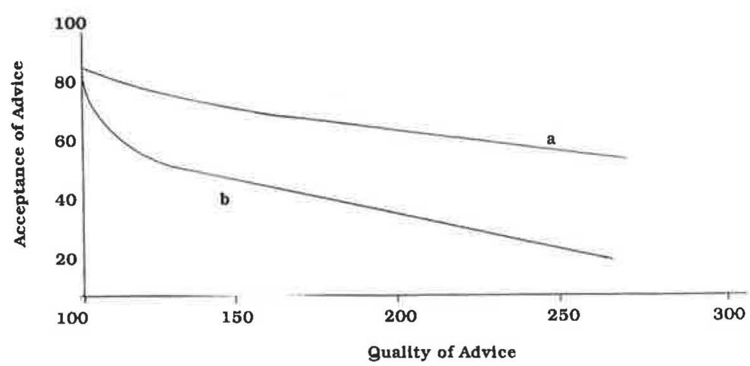
Apparently participants who had become accustomed to receiving good advice became less critical of the occasional bad piece of advice, either because they did not feel it necessary to question it or because they were inclined to give it the benefit of the doubt. On the other hand, participants who have been experiencing a lot of bad advice seemed to treat any advice with great skepticism.

Further analysis, not reproduced here, suggests that the average quality of the most recently received items of advice was particularly influential in establishing the credibility, or otherwise, of the current advice. The data was also examined to see if any primacy effect existed—whether the quality of the first few items of advice was particularly important. No such effect was apparent.

Acceptance as a Function of Familiarity with the Network

Acceptance of advice generally decreased as familiarity with the network increased. Table 1 shows that, among people who had been receiving fairly reliable advice, it fell in a decreasing curve, dropping by about 10 percent between the first and second journey in a given network, by about 50 percent between the second and third journey, by about 2 percent between the third and fourth, and by about 1 percent per journey thereafter. The decline in acceptance among people who had been receiving poor advice was much less regular and appeared to be very dependent on actual conditions met.

Acceptance was highest when a new destination had to be found in a new network for which no map was available; in such circumstances even those who had been receiving very poor advice had little option but to rely on it. Acceptance was lowest when a journey was being made in a network for



a by participants whose previous advice has provided routes averaging within 3% of the theoretical minimum journey time

b by participants whose previous advice had provided routes averaging at least 20% longer than the theoretical minimum journey time.

FIGURE 4 Acceptance of advice as a function of the quality of previously received advice.

TABLE 1 ACCEPTANCE OF ADVICE AS A FUNCTION OF FAMILIARITY WITH THE NETWORK

Journey Nº	Nº of previous journeys in this network	Nº of previous visits to this destination	Map available	% of advice accepted a	b	c
4	3	2	Y	79	24	72
5	4	3	Y	77	56	70
6	5	1	Y	85	56	69
7	6	4	Y	76	59	67
8	7	5	Y	75	65	68
9	8	2	Y	81	58	68
10	0	0	N	96	79	83
11	1	1	N	84	73	77
12	2	0	N	89	74	76

a when previous advice had provided routes averaging within 3% of the theoretical minimum journey time

b when previous advice had provided routes averaging at least 20% longer than the theoretical minimum journey time

c for all qualities of previous advice.

Source: DRIVE V1011 tests with IGOR in UK and France

which a map was available, to a destination that had been visited several times before.

Although acceptance of advice declined with increasing familiarity, participants' adherence to preplanned routes increased as their confidence grew.

The Effect on Acceptance of Corroborating or Contradictory Evidence

The influence of circumstantial evidence that tended to corroborate or contradict the advice was studied in some detail. The results reported in Table 2 show that some features—particularly the alignment of the advised exit relative to the crow-fly direction to the destination, the behavior of other drivers, and the presence or absence of congestion—had a significant impact.

The impact was particularly strong when the advice itself was not optimal (i.e., when the advised exit was not the one that would have got participants most quickly to their destinations). Thus, if non-optimal advice happened to be in the right direction (in terms of compass bearing) it was accepted by 74 percent of participants but if it was in completely the wrong direction it was accepted by only 22 percent of participants. Similarly, if the non-optimally advised exit happened to be used by most other drivers, then the advice was accepted by 67 percent of participants, but if it was the least used exit by other drivers it was accepted by only 30 percent of participants. The visible presence of traffic congestion on all exits other than the advised one, or of a road sign apparently confirming the advice, also has an important effect on the acceptance of non-optimal advice.

Acceptance of optimal advice was not influenced by corroborating or contradictory evidence to quite the same extent as was acceptance of non-optimal advice. Even so, the effect of compass direction, visible congestion, and the behavior of other drivers was very important.

TABLE 2 ACCEPTANCE OF ADVICE AS A FUNCTION OF THE EXTENT OF CORROBORATION OR CONFLICT FROM OTHER SOURCES OF INFORMATION

Source of corroboration or conflict	% of Optimal advice accepted with: (a) Max corrob-oration	(b) Max conflict	% of Non optimal advice accepted with: (a) Max corrob-oration	(b) Max conflict
Crowfly direction indicator	91	56	74	22
Other drivers' decisions	85	76	67	30
Traffic congestion visible on exits	90	74	58	44
Signposts	80	80	59	42
Size of road	80	80	49	52

Source: DRIVE V1011 tests with IGOR in UK

It appears reasonable to conclude from the evidence presented in Table 2 that participants' reaction to advice was strongly conditioned by local network conditions and their perception of the physical layout of the network. To the extent that these are correlated with optimal advice, this no doubt explains how the acceptance comes to be so closely related to the quality of guidance.

The SP exercise conducted with each participant immediately after their last journey with IGOR allowed the further exploration of the effect of corroborating and contradictory evidence on acceptance and rejection of advice.

The SP results were similar to those from the main IGOR exercise, in as much as they showed the importance of congestion and of the crow-fly alignment and the relative unimportance of road size and signposting, but there was some evidence that the fixed ordering of questions in the SP experiment had influenced the results. Problems such as this frequently occur in conventional questionnaires but are avoided in the main IGOR exercise because the order and content of situations faced by participants vary from participant to participant and from journey to journey.

Acceptance of Advice as a Function of Personal Characteristics

There was some relationship between participants' apparent propensity to accept route-guidance advice and their personal characteristics. In many cases, however, the relationship was not significant at the 5 percent level and in some cases quite different relationships were apparent in the British and French samples. Among those relationships that were significant at the 5 percent level in the British sample, women were found less likely to accept advice, particularly non-optimal advice, than men (38 percent compared with 46 percent of non-optimal advice was accepted, respectively) and people who have a high annual kilometrage or who regularly drive to work were less likely to accept advice than others. It was also noted that acceptance of advice, particularly non-optimal advice, was highest amongst people who quoted distance minimization as their main criterion for route choice (71 percent of non-optimal advice accepted compared with 47 percent for the whole sample). Also, those who said they had previously heard of route guidance, and had thought it likely to be useful to them for most of their journeys, accepted advice more readily than the population at large (52 percent compared with 47 percent).

DISCUSSION OF RESULTS

Practicality of the Tool

Experience with IGOR in the United Kingdom and France has confirmed the expectation that it would provide a cost-effective method of collecting large amounts of data. The management and control of the surveys, the handling of the data, and the speed of subsequent data processing have all proved very satisfactory (though not fool proof—almost 10 percent of the data set was lost because of a disk error on the French machine).

IGOR appears to have been popular with participants and no difficulties were experienced either with recruitment or with people wishing to terminate a session part way through (the average session lasted about 35 to 40 min and respondent fatigue might have been expected to set in had it been a conventional questionnaire or interview).

Although IGOR is self-contained and could in theory be run without any survey staff in attendance, in practice it was found useful to have someone present to record any comments made by the participant at any stage during the session and to seek more attitudinal information from them at the end of the session. Comments made by participants while they were deciding which exit to take proved very revealing as to participants' decision processes and are perhaps a valuable source of data in their own right.

The idea of mailing IGOR disks to people with access to their own PC's had little success because of the current concern about the importation of computer viruses by means of direct-mail disks.

Reliability of the Results

IGOR puts the participants in a simulated route-choice situation and provides them with simulated guidance but neither of these is the real thing. Participants do not receive real environmental stimuli and do not work within the same constraints as they would were they making real journies. The stimuli could be improved albeit at the cost of reduced portability, by use of more sophisticated graphics or simulators, such as is being done by research teams interested in the ergonomic aspects of IVRG (*15–18*) but even so, the situation would still be artificial.

Having conducted some 350 IGOR sessions, followed by debriefing sessions with each of the participants, there is confidence that they understood what they were doing and were interpreting the on-screen information (such as that relating to the decisions of other drivers and the disposition of signposts) realistically. There are, however, two reservations; one relates to the crow-fly direction indicator and the other to the way in which feedback on performance was given.

The crow-fly indicator was supposed to represent the drivers' general sense of direction and it probably fulfilled this task quite well for journeys made in a familiar network or with a map. However, given that only a minority of drivers carry directional compasses in their vehicles, it may be that it was unrealistically precise for the journeys being made in an unknown network without a map. If this is so it will probably have depressed the acceptance of advice on such journeys lower than what it might otherwise be; the consequence of which would be that the acceptance would be even more sensitive to familiarity than the results suggest. In future versions of IGOR, the intention is to experiment with less precise direction indicators and, for some journeys, with having no direction indicator at all.

At the end of each journey participants were reminded of how long they might have expected the journey to take and informed of how long they actually did take. This information was intended to give them some fairly realistic feedback on their performance. There is, however, some evidence, particularly from among the French participants, that the information was treated as a score and may have encouraged some participants to seek minimum time routes to a greater extent than they would do in real life (only 35 percent of French participants quoted time minimization as their overriding criteria for route choice on the journey to work). If this is so it may have caused the relationships between acceptance of an item of advice and its (objective) quality to be stronger than they might otherwise be. In future versions of IGOR other forms of feedback will be experienced with, including measures tailored to an individual participant's stated route-choice criteria.

The analysis suggested that personal characteristics generally had influence on an individual's acceptance of advice and the authors are confident that, with the possible exception of the points outlined above, the results are probably a fairly good indicator of how people might react to route guidance in real life. It would, of course, be nice to be able to confirm this by comparing IGOR's results with observations from real schemes. The problem, of course, is that there is insufficient evidence from real schemes against which to judge the IGOR results (indeed had such evidence existed, IGOR would not have been necessary). The IGOR results are, however, consistent with those obtained from attitude surveys among drivers equipped with Ali-Scout equipment in Berlin's LISB system (*2,3*). More rigorous tests of the IGOR results must await the availability of large volumes of data on individual drivers' responses such as might be obtained by automatic monitoring of vehicle movements as part of a carefully designed field trial.

Until such evidence becomes available, IGOR is surely one of the best sources of quantified data on which to base models of drivers' response to route-guidance advice. It is certainly to be preferred to stated intentions, preferences, and attitudes derived by more conventional means.

Further Work

Unless equipped drivers form only a trivial proportion of the driving population, their behavior, whether or not it is in compliance with advice, could materially affect network conditions. A realistic model of route guidance or information systems must incorporate a representation of mechanisms on the demand-side as well as on the supply-side. Much might be gained by embedding the calibrated models of driver response, derived from IGOR, within one of the network simulation modeling frameworks currently being used to examine driver information systems (*19–24*).

IGOR has proved to be a very valuable tool with which to examine route choice. It has already enabled experiments with a range of situations that cannot readily be observed in the field. A natural progression appears to be to use it to study a range of route-choice issues and to represent a variety of driver information systems. Plans are underway to represent systems that provide text or map-display information as well as, or instead of, guidance.

The process of developing IGOR itself highlighted a number of interesting issues relating to route-choice behavior—for example the role of feedback on performance, the role of pretrip planning, and the accuracy of individuals' knowledge of their orientation. We intend to examine these and other issues in future versions of IGOR (*25*).

Analysis of the data collected in Britain and France during 1990, and of comments by participants and survey staff, has raised important issues, some of which will benefit from further analysis of the existing dataset, whereas others will require further experimentation. Further analysis will, it is thought, throw more light on the way in which different factors (e.g., qualities of guidance and corroborating evidence of various kinds) act in combination to influence acceptance of advice, whereas exploration of participants' motivations will require further experimentation.

The use of IGOR in Britain and France during 1990 was supported under the DRIVE initiative in order to assist in the prediction of the extent to which drivers will accept route-guidance advice. Plans are underway, as part of an SERC sponsored project at the Universities of Leeds, Southampton and York, to generalize the program, to look more broadly at route choice and network learning behavior. It is hoped that IGOR will be able to produce data to support the specification of more general route-choice models in which guidance is merely one of several potential explanatory variables.

ACKNOWLEDGMENTS

The development of IGOR was supported by a grant from the UK Science and Engineering Research Council. We are grateful to colleagues at ITS, particularly Jeremy Pacht and Matthew Joint for their comments on the original concept and early versions of the program and to anonymous referees for their comments on earlier versions of this paper. The data presented in the paper was collected by ITS and INRETS within the CARGOES consortium under the European DRIVE initiative. The interpretation of the results and the views expressed must, however, remain the responsibility of the authors.

REFERENCES

1. P. W. Bonsall and T. Parry. Drivers' Requirements for Route Guidance. *Proc., 3rd International Conference on Road Traffic Control*, IEE, London, May 1990.
2. R. Slapa and P. W. Bonsall. *Questionnaire Survey of LISB Users' Route Choice Behavior and Response to Guidance*. ITS WP295, Leeds University, 1990.
3. M. Joint and P. W. Bonsall. *Questionnaire Survey Among Users of the LISB Dynamic Route Guidance System*. ITS WP321, Leeds University, Feb. 1991.
4. *Drivers' Reaction to Accuracy of Guidance Information*. CARGOES, Deliverable 20 of Project V1011, Siemens, Munich, Nov. 1990.
5. P. W. Bonsall, L. Pickup, and A. Stathopoulos. Measuring Behavioural Responses to Road Transport Informatics. *Proc. DRIVE Conference*, Vol. 2, DRIVE Office, Brussels, Feb. 1991, pp. 1457–1487.
6. A. D. May and P. W. Bonsall, *Recommended Methodology for the Evaluation of Route Guidance in London and the Potential for Comparability with Berlin*. CR130, Department of Transport, TRRL, Crowthorne, England.
7. M. A. Bradley, P. M. Jones, and E. Ampt. An Interactive Interview Method to Study Bus Provision Policies. *Proc., 15th PTRC Summer Annual Conference*, Vol. C, PTRC, London, 1987, pp. 147–162.
8. M. A. Bradley, T. Grosvenor, and A. Bouma. An Application of Computer-Based Stated Preference to Study Mode Switching in the Netherlands. *Proc., 16th PTRC Summer Annual Conference*, Vol. D, PTRC London, 1988, pp. 307–319.
9. P. M. Jones, M. C. Dix, M. I. Clarke, and I. G. Heggie. *Understanding Travel Behaviour*. Oxford Studies in Transport, Gower, Aldershot, 1983.
10. P. W. Bonsall. *A Survey of Attitudes to Car Sharing: A Database for Micro Simulation*. Department of Transport, TRRL SR563, TRRL, Crowthorne, England 1980.
11. H. S. Mahmassani and G.-L. Chang. Experiments with Departure Time Choice Dynamics of Urban Commuters. *Transportation Research 20B(4)*, 1986, pp. 297–320.

12. H. S. Mahmassani, G.-L. Chang, and R. Herman. Individual Decisions and Collective Effects in a Simulated Traffic System. *Transportation Science 20(4)*, 1986, pp. 258–271.
13. H. S. Mahmassani and R. Herman. Interactive Experiments for the Study of Tripmaker Behaviour Dynamics in Congested Commuting Systems. *Oxford Conference on Travel Behaviour—New Developments in Dynamic and Activity-Based Approaches*, Oxford, England, 1988.
14. P. W. Bonsall and T. Parry. A Computer Simulation Game to Determine Drivers' Reactions to Route Guidance Advice. *Proc., 18th PTRC Summer Annual Conference*, Vol. H, PTRC London, England, 1990, pp. 113–124.
15. P. M. Jovanis and R. Kitamura. User Perceptions and Safety Implications of In-Vehicle Navigation Systems. *Proc. 1st VNIS Conference*, Toronto, Ontario, Canada, Sept., IEEE, 1989, pp. 113–116.
16. B. W. Stephens. Comparison of Alternative Methods for Presenting Trip Navigation Information to Motorists. *Proc., 18th PTRC Summer Annual Conference*, Vol. G, PTRC, London, England, 1990.
17. R. Van der Horst and W. Janssen. *Measuring Behavioural Effects of Dynamic Route Information Systems*, DARTS Working Paper, Institute for Perception, TNO Delft, The Netherlands, 1990.
18. N. Ayland. Evaluating Real Time Responses to In-Vehicle Driver Information Systems. *Proc., DRIVE Conference*, Vol. 1, DRIVE Office, Brussels, Feb. 1991, pp. 70–88.
19. H. S. Mahmassani and R. Herman. System Performance and User Response under Real Time Information in a Congested Traffic Corridor. Presented at seminar on Urban Traffic Networks, Capri, Italy, June 1989.
20. J. C. Smith and K. Russam. Some Possible Effects of Autoguide on Traffic in London. *Proc., 1st VNIS Conference*, IEEE, Toronto, Ontario, Canada, Sept. 1989, pp. 282–287.
21. T. Van Vuren, D. Van Vliet, and M. J. Smith. Combined Equilibrium in a Network with Partial Route Guidance. Presented at the Engineering Foundation Conference on Traffic Control Methods, Santa Barbara, Calif., 1989.
22. L. Breheret, N. B. Housell, and M. McDonald. The Simulation of Route Guidance and Traffic Incidents. Paper presented at UTSG Conference, Hatfield Polytechnic, United Kingdom, Jan. 1990.
23. M. W. Van Aerde and S. Yagar. Combining Traffic Management and Driver Information in Integrated Traffic Network. *Proc., 3rd International Conference on Road Traffic Control*, CP320, IEE, London, England, May 1990, pp. 11–15.
24. P. W. Bonsall. The Influence on Route Guidance Advice on Route Choice in Urban Networks. *Proc. Jap. Soc. Civ. Eng.* 425/IV-14 1991-1.
25. P. W. Bonsall. IGOR—Further Work Planned Using the Route Choice Simulator. ITS TN290, University of Leeds, United Kingdom, 1991.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

Comparative Assessment of Origin-Based and En Route Real-Time Information Under Alternative User Behavior Rules

HANI S. MAHMASSANI AND PETER SHEN-TE CHEN

The effects of real-time traffic information, supplied at the origin of the trip or along the way (en route), on the system's performance under alternative behavioral rules governing path selection in the network are examined in this paper. Simulation experiments are performed to investigate the effect on overall system performance as well as the incidence of benefits (costs) across user information groups of four experimental factors: (*a*) behavioral rules, governing users' response to real-time information, (*b*) sources of information, consisting of point-of-departure or in-vehicle (or both), (*c*) prevailing "initial conditions" in the system, and (*d*) market penetration (i.e., the fraction of users with access to real-time information in the network). The results of these simulation experiments provide insights into the effectiveness of real-time advanced driver information systems on systemwide performance and on its critical determinants. The results confirm a priori expectations that the existence of benefits as well as the relative effectiveness of origin-based versus en route information is highly dependent on the initial conditions prevailing in the system as well as the behavioral rules governing path selection. Extreme behavior by users (with frequent switching in myopic response to any gain, no matter how small) could lead to severe worsening of traffic conditions under real-time information from either source. On the other hand, switching according to a boundedly rational model incorporating a threshold improvement in trip time is more likely to lead to meaningful systemwide benefits.

The supply of real-time traffic information to tripmakers through various possible forms of advanced driver information systems (ADIS) is increasingly looked to as a means of reducing traffic congestion in urban networks (*1*). The effectiveness of different types of ADIS and alternative information supply strategies depends on the complex interactions between user decisions and the network's performance. Questions regarding user behavior as it pertains to different types of ADIS do not appear to have received much attention to date, although they will be critical to the successful implementation of these systems.

In this paper, concern is with two principal behavior aspects of real-time ADIS. The first consists of the mechanism or choice rule governing path selection in response to real-time information on the current trip times on competing paths. The second concerns market penetration, or the effect of the fraction of users in the network with access to such information. While models are proposed and illustrated for the first phenomenon, with respect to the latter, sensitivity analysis is performed and overall system performance and effectiveness is examined for different levels of market penetration.

Department of Civil Engineering, The University of Texas at Austin, Austin, Tex. 78712.

Another principal focus of this paper is the comparison between two sources of information with different technological and cost characteristics: point-of-departure information versus in-vehicle (en route) information. In both cases, the user would have access to real-time information on current conditions on the various network components. However, in the former, the information influences the driver's initial choice of route (and possibly his or her time of departure), whereas in the second case the driver is already on an initially selected route after having left his or her origin. The analysis focuses primarily on the a.m. home-to-work commute (though a similar treatment could be developed for the p.m. peak period). The point-of-origin information then becomes equivalent to information that one would consult at home, and which can therefore be supplied through existing media, primarily television or telephone lines (connected to a home terminal). On the other hand, in-vehicle information would require the kind of telecommunications and microprocessor technologies and infrastructure contemplated in the various proposed and prototype ADIS systems (*1*). The relative effectiveness of these two sources of traffic information has extremely important technical and policy implications, given the significant differences in the size of the respective investments required by each source. In other words, which of the two sources, taken separately, leads to greater improvement in system performance, if any? If origin-based information is provided, how much additional improvement could be attained by continued en route information availability?

The above questions are explored in the context of numerical simulation experiments in a commuting corridor. Several key factors that determine the answers to these questions are identified and illustrated, particularly the sensitivity of the results to (*a*) the underlying behavioral rules governing users response to the supplied information, (*b*) market penetration, or the fraction of the population with access to the information, and (*c*) the prevailing "initial conditions" in the system, and the corresponding availability of improvement opportunities via flow redistribution among alternate routes.

The simulation experiments are performed using a modeling framework previously developed by Mahmassani and Jayakrishnan (*2*). A corridor network structure is considered, consisting of three principal highway facilities offering several switching opportunities. Only auto drivers are considered, though the behavioral framework could be extended to include transit users. As noted, the illustration is limited to the a.m. home-to-work commute.

In the next section, some aspects of the behavioral framework for the user response models are discussed, and specific decision rules are formulated for this purpose. The simulation experiments are then described, including an overview of the simulation model used and a detailed discussion of the specific levels considered for each experimental factor. The simulation results are then discussed, followed by concluding comments.

BEHAVIORAL FRAMEWORK AND RULES

General Behavioral Considerations

The effectiveness of information technology to achieve traffic control objectives depends on (*a*) the existence of improvement opportunities in prevailing traffic conditions, (*b*) the nature and type of information available to different segments of the user population, and (*c*) users' behavior and response to the supplied information.

Mahmassani and Jayakrishnan (*2*) have classified information strategies into four generic categories:

1. Descriptive, stored information, such as a static map that displays only stored information on (fixed- or time-dependent) trip times on the various network links.
2. Descriptive, real-time information, where the trip times are updated on a real-time basis to indicate prevailing congestion on the various network links.
3. Descriptive, real-time information with individual optimization, in which the link-level information could be processed either on board or centrally to compute the current shortest path from the present position to the desired destination of given driver.
4. Controlled guidance, under which the instructions given to users reflect a central controller's system-level objectives, subject to certain constraints to prevent unreasonable penalties to any individual tripmaker.

User behavior and response to the supplied information is the result of a complex process involving human judgment, learning, and decision making in a dynamic environment. This process depends on the type and nature of the information provided, in addition to the individual characteristics and preferences of the tripmaker. In all four cases identified above, user response can be viewed in terms of four choice dimensions: (*a*) acquisition of the equipment, (*b*) consultation of the information system, (*c*) compliance with its instructions, and (*d*) trip decisions and actions (behavior). Of course, these dimensions take place over varying time frames.

Acquisition of the on-board equipment is a long-term decision, reached by the user after some deliberation involving the trade-off of perceived benefits with monetary costs. This decision can undoubtedly be modeled adequately in a standard random utility maximization discrete choice framework.

Consultation of the equipment is a real-time decision, made along and possibly also at the beginning of the trip. The consultation decision is likely to be governed by different behavioral mechanisms for each of the preceding generic types of information systems and strategies. For example, a static map with historic information only (the first strategy) is not likely to be consulted (for path selection purposes) other than at the beginning of a trip or to find an alternate path along the way if actual congestion exceeds anticipated congestion. [Operationally, this choice can be modeled as follows. Let TT_{in} denote the actual trip time experienced by user i up to node n, and ATT_{in} the initially anticipated trip time up to that node. If $(TT_{in} - ATT_{in})$ exceeds a certain threshold, then the driver will consult the video map to identify a suitable alternative.] On the other hand, a controlled guidance system (the fourth strategy) is likely to be consulted on a virtually continuous basis. The underlying behavioral process for the second strategy is likely to be closer to that for the first strategy, whereas the third strategy is likely to be closer to the controlled guidance case. In all cases, the consultation decision is influenced by longer-term processes, particularly user perceptions of the reliability and usefulness of the system, formed mostly by learning through one's own experience with the system, as well as reports by friends, colleagues, and popular media.

The third choice dimension is referred to here as compliance. It is a real-time decision, definitely influenced by the above-mentioned longer-term learning phenomena that form user perceptions. This dimension is not directly applicable to the first two information supply strategies. It is most applicable in the fourth strategy, where specific route guidance instructions are provided, with the controller's intent that they be followed by the motorists. In this case, compliance will be a critical factor in overall system effectiveness. For the third strategy, the applicability of this dimension depends on the specific information displayed. If it consists of just (current) trip times on alternate paths, then it is not clear what compliance would refer to. On the other hand, if a single path is displayed, because it has been determined to be the shortest or otherwise "best" according to some criterion (or as a result of an on-board expert system recommendation), then compliance could be defined relative to that recommendation.

The fourth dimension actually includes several possible decisions, with varying time frames for each. The most evident is en route path switching, which is the principal real-time decision targeted by in-vehicle information systems. A second decision consists of initial (home based for the a.m. commute) route selection. This decision can be taken as an immediate response to real-time information consulted at the trip origin (and supplied either through the in-vehicle display or some other in-home medium). However, it is also influenced by the day-to-day experience of the users in the system, which contributes to forming the users' perceptions both over the short-term (day-to-day) and over a longer time frame. A third response consists of trip timing. In the immediate term, the tripmakers can delay, advance, or both the time of their trips, in light of prevailing traffic conditions. From day-to-day, the tripmakers will adjust their departure time, as a result of learning through repeated system usage. The user's preferred arrival time plays an important role in this process. Some equilibrium choices or timing strategies might be reached by individual commuters over time as a result of this process. (Another possible choice available to users in the medium term is that of changing modes, particularly switching from automobile to transit, if adequately available.) Over the medium to long term, changes in activity patterns could take place, reflecting the users' potentially improved ability to schedule their activities using reliable real-time information. Of course, over the very long term, the standard choice dimensions of residential or work locations remain available for users; these are outside the scope of the present discussion.

An important factor that was noted in connection with all of the decisions just outlined is that of the perceived reliability of the information and its resulting credibility. This arises primarily from the dynamic nature of the decision environment and the presence of collective effects in the network as a result of the interactions of a large number of individual decisions. In particular, a "best" path predicated on current link-trip times may well turn out to be less than optimal as congestion in the system evolves. This issue is of particular concern for the third and fourth information strategies defined previously. Of course, it is possible to use predicted trip times for path computations instead of the current actual values. However, the accuracy of the prediction logic would have to be questioned as no such entirely satisfactory prediction techniques are currently available. This is particularly problematic under the fourth strategy (controlled guidance). To what extent will (and can) the anticipated response of users to the supplied guidance instructions be incorporated in the prediction? Naturally, the reliability issue becomes more critical as the fraction of users in the system with access to the information increases, as seen in the simulation experiments presented in the final section of the paper.

The focus of this paper is primarily on the en route path switching decision in real-time, in response to information of the third strategy. In other words, the users are assumed to know the current travel time on the best path from the current node to their destination. The focus is also on the initial path selection decision, at the trip origin, in response to information similar to that available under the third strategy, though it may not necessarily be delivered through an in-vehicle medium, as discussed in the previous section. Only the real-time aspects of these decisions are considered in this analysis; the day-to-day and longer-term aspects will be introduced at a later stage. In addition, the equipment acquisition decision will be indirectly considered by conducting sensitivity analyses with respect to a market penetration parameter.

Next, alternative behavioral mechanisms and choice rules for path switching and initial route selection are formulated. These form the basis of the numerical simulation experiments conducted to compare the relative effectiveness of en route and origin-based real-time information.

Path Selection and Switching Rules

Several alternative mechanisms and behavioral rules with varying degrees of complexity could be formulated for each of the elements identified in the preceding discussion. However, in the absence of definitive observational data, the focus is on simple operational representations that adequately capture the character of user behavior and can thus be used in the context of simulation models to evaluate the overall performance and effectiveness of real-time information in traffic networks. Three distinct alternative rules are proposed hereafter for en route path switching; similar constructs can also be used for initial path selection. The first is a so-called myopic deterministic choice rule; the second is a boundedly rational model of path switching, whereas the third is a standard random utility discrete choice model. These are described in turn hereafter.

Rule R.1. Myopic Switching Rule

By analogy to the standard path choice rule in static network assignment models (*3*), this rule simply states that from any given node n, users will always select the best path (in terms of least cost or least travel time) from the current node to their destination, that is,

$$\delta_{ik}(n) = \begin{cases} 1 \text{ if } TT_{ik}(n) \le TT_{i}(n), \ \forall\, l \in P_i(n),\ i \neq k \\ 0 \text{ otherwise} \end{cases}$$

where

$\delta_{ik}(n)$ = binary indicator, equal to 1 if user i selects path k between node n and the destination, and 0 otherwise;
$TT_{ik}(n)$ = current trip time on path k between node n and user i's destination; and
$P_i(n)$ = set of paths between node n and user i's destination.

The assumption that drivers follow the current best path from every node along the way is rather extreme, in that it would lead the user to switch paths in pursuit of any gain, no matter how insignificant.

Rule R.1 can be restated in the following switching form:

$$\delta_i(n) = \begin{cases} 1 \text{ if } TTC_i(n) > TTB_i(n) \\ 0 \text{ otherwise} \end{cases}$$

where

$\delta_i(n)$ = binary indicator equal to 1 if user i switches from the current path to the "best" one between node n and the destination, 0 otherwise;
$TTC_i(n)$ = trip time on the current path from node n to user i's destination; and
$TTB_i(n)$ = trip time on the best path between n and user i's destination.

An important concept in this rule is the notion of a current path, which is central to the modeling framework. It assumes that the users have an evoked current path to which they might exhibit some degree of commitment. In a freeway corridor context, such an evoked path might be strongly associated with the freeway itself or with a major alternate parallel arterial. A possibly more reasonable assumption than the above extreme myopic rule is that driver switching behavior exhibits a boundedly rational character anchored in one's current path. This assumption is operationalized next.

Rule R.2. Boundedly Rational Switching Rule

The notion of bounded rationality in the context of the day-to-day dynamics of departure time and route decisions of urban commuters has been extensively explored by Mahmassani and Chang (*4–7*). It has been applied for the real-time route switching decisions under in-vehicle information by Mahmassani and Jayakrishnan (*2*). This notion can be operationalized using a satisficing switching rule with an indifference band of trip time saving. Users will switch from their current path to the best alternative only if the improvement in the remaining trip time exceeds this indifference band. This threshold can be expressed either in absolute terms, or relative to the remaining trip time. Following Mahmassani and Jayakrishnan (2), a switching rule with a relative indifference band subject to a minimum (absolute) trip time saving can be stated as

$$\delta_i(n) = \begin{cases} 1 \text{ if } TTC_i(n) - TTB_i(n) > \max\,[\eta_i(n) \cdot TTC_i(n), \tau_i(n)] \\ 0 \text{ otherwise} \end{cases}$$

where

$\eta_i(n)$ = relative indifference band for user i, as a fraction of the remaining trip time on the current path from node n to the destination: $TTC_i(n)$, with $\eta_i(n) \geq 0$, $\forall\, i, n$;

$\tau_i(n)$ = minimum improvement in the remaining trip time, from node n to the destination, necessary for user i to switch from his or her current path, with $\tau_i(n) \geq 0$, $\forall\, i, n$; and all other terms are as previously defined.

It can be seen that rule R.1 is a special case of rule R.2 with $\eta_i(n) = 0$ and $\tau_i(n) = 0$, $\forall\, i, n$.

In the model, $\eta_i(n)$ is expressed in relative terms. It can be thought of as the percent improvement in remaining trip time vis-a-vis the current path. Moreover, to preserve a meaningful threshold effect and preclude unintended switching when $TTC_i(n)$ becomes very small as drivers approach their destination, the absolute band $\tau_i(n)$ is introduced to provide a lower bound. Both $\eta_i(n)$ and $\tau_i(n)$ could be either fixed constants or vary from node to node, and possibly over time. Furthermore, they could be related systematically to the socio-demographic attributes of the user. The simulation results presented in this paper assume fixed values for these bands over the duration of any given trip. Furthermore, while $\eta_i(n)$ is allowed to vary across users, $\tau_i(n)$ is taken as constant τ for all drivers.

Several alternative formulations for the indifference bands are possible. For instance, the band could increase at a fixed percentage rate per distance traveled, or according to some systematic process that would capture the dynamic effect of the user's experience in the system. Such increasing bands would reflect the drivers' decreasing propensity to switch as they approached their destinations.

Rule R.3. Probabilistic Discrete Choice Rule

A natural alternative mechanism for path selection at a given node would be a probabilistic discrete choice model, along the lines of those proposed for stochastic network traffic assignment (8). For example, the well known multinomial logit form could be used to calculate the probability $P_{i,n}(r|R_n)$ that user i chooses route r from node n to the destination, given a choice set R_n of alternative paths available at that node. This probability would be a function of the utilities $U_{in}(r)$ associated with each route $r \in R_n$. The utilities would be functions of the characteristics of the routes, including the trip time, trip time reliability, schedule delay, and a possible penalty for switching from the current path. Some difficulties do, however, arise in the application of such a route choice rule for repeated en route decisions in the context of a simulation model. The principal concern is that such decisions cannot be assumed to be independent within and across users. Specifications of the utility functions in a way that correctly capture the temporal dependencies within each user as well as across users are likely to be quite cumbersome, and rather difficult to calibrate correctly and implement operationally.

As noted previously, the rules could be applied en route as well as at the trip origin, primarily in connection with descriptive real-time information with self-optimization capability (information supply strategy three), which would provide estimates of the remaining trip time on the user's current path as well as identify the best path (for rules R.1 and R.2). For rule R.3, additional information would be required, though its form and content is not entirely clear. An expert system could possibly provide trip times on a subset of efficient paths, including the current one. All of the above rules can be modified for multiple destinations by adding the appropriate destination subscript.

Departure Time Decision Considerations

The preceding rules are for the path selection and switching decisions, which form the focus of the experiments reported in this paper. Rules for the departure-time decision in response to real-time information have not been investigated nor suggested by researchers, as the effect of real-time information on this choice dimension has not generally been recognized. However, simulation results reported by Mahmassani and Jayakrishnan (2) for a particular set of conditions in a hypothetical commuting corridor have suggested that the potential to reduce congestion through peak spreading obtained by trip-time shifting may be quite significant relative to what might be achievable through traffic redistribution over space (i.e., over alternate paths). This point has also been highlighted in a recent Mobility 2000 report (9).

As discussed previously, departure time changes could take place both in real-time as well as from day-to-day (and eventually over the long term). On a real-time basis, the user can consult a home-based advisory unit (or TV-relayed information) for current traffic conditions, and decide accordingly to advance, delay, or keep the trip at the same time. Rules with varying degrees of complexity could be formulated for this decision process. Two classes of rules can be suggested for this purpose: (*a*) satisficing rules, that trigger a change when the difference between anticipated and current conditions exceed some threshold; and (*b*) random utility maximization rules whereby users select the departure time that maximizes their utility over all subsequent possible departure times [examples of specifications for the utility functions can be found in Hendrickson and Plank (*10*)].

Of greater concern are the day-to-day adjustments of departure time and the resulting longer-term changes in the temporal loading pattern. Rules similar to those proposed by Mahmassani and Chang (*11,12*) and Tong, et al. (*13*) could be adapted to the situation of real-time information availability. Essentially, such information would be combined with the user's experience with the network to provide the basis for such departure time adjustments (*12,13*). With such rules, the evolution of the system under real-time information could be explored, and the effectiveness and benefits of the latter could be assessed from a long-term perspective rather than simply through the consideration of real-time responses on a given day. Such investigation is outside the scope of the present paper, which considers only real-time path selection responses of commuters. The simulation-assignment modeling framework and the specific simulation experiments conducted for this study are described in the next section.

MODELING FRAMEWORK AND SIMULATION EXPERIMENTS

This section first describes the commuting context and some assumptions pertaining to the definition of alternate paths in the network, followed by an overview of the simulation-assignment methodology. The simulation experiments are then described, including the four principal experimental factors and the corresponding levels considered.

Commuting Context

The model used to perform the simulation experiments is an extension of the corridor simulation-assignment model developed by Mahmassani and Jayakrishnan (*2*). In addition to the existing en route path switching model, it incorporates a new pretrip path selection component to evaluate system improvement opportunities under real-time on-board information, real-time home-based information, or both. The simulation experiments are performed for a commuting corridor with three major parallel facilities, such as freeways or major arterials, for the morning work commute. For convenience and with no loss of generality, all three facilities are 9-mi long, and each is discretized into 9 segments, each 1-mi in length, with crossover links at the end of the 3rd, 4th, 5th, and 6th mi to allow switching from one facility to another (see Figure 1). Commuters enter the corridor through ramps feeding into each of the first six (1-mi) segments on each facility, and commute to a single common destination downstream (such as the central business district or a major industrial park).

In all experiments, 10,800 vehicles are loaded to the corridor over a total duration of 75 min according to a specified time-dependent departure pattern (which is itself an experimental factor in the simulations, as described later in this section). Of the three major facilities, hereafter referred to as Highways 1, 2, and 3, Highway 1 has the highest free mean speed of 55 mph. Highway 2 is the second best, with a free mean speed of 45 mph, followed by Highway 3 (35 mph). All the crossover links have a free mean speed of 45 mph.

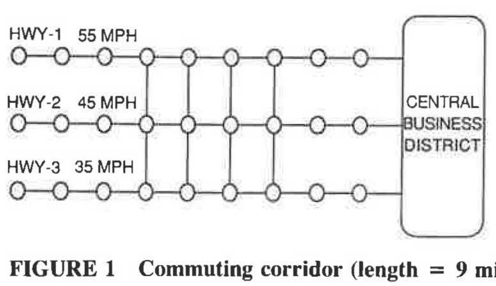


FIGURE 1 Commuting corridor (length = 9 mi).

Assume that a fraction of the commuters have access to real-time network traffic information, be it from an on-board traffic advisory unit or a home-based traffic advisory unit. In particular, the user receives information on the prevailing trip times on all the links of the network. These form the basis for computing the trip times from the user's present location (either at the origin or en route) to his or her destination along alternate paths (i.e., under the third information supply strategy discussed in Section 2). A behavioral assumption is made in the definition of available paths, namely that users perceive and identify a path in terms of its major highway facility. This would result from the extent of a hierarchy in the manner in which users perceive a particular network. Thus a path for the purpose of this analysis consists of a single major facility (to the destination) along with its connecting link. Consequently, at any given node (including the origin), the user effectively considers only three paths, one for each facility. In addition to its behavioral plausibility, this assumption offers the advantage of tremendously simplifying the path processing burden that must be performed by the assignment model (*2*).

To implement the behavioral rules presented in the preceding section (particularly rules R.1 and R.2), a current path is associated with each user, as discussed previously, both at the origin node and along the way. In the latter case, an alternative route consists of a crossover link from the current node to another major highway facility and the remaining portion of that facility to the destination. Thus at every node (including the origin), $TTC_i(n)$ and $TTB_i(n)$, the travel times on the current and the best paths, respectively, are calculated for processing by the user decisions component, as explained in the following.

Simulation-Assignment Model

The model is composed of three main components: the traffic performance simulator, the network path search and assignment component, and the user decision-making component. The first component is a special-purpose macroparticle traffic simulator (MPSM), which extends a code previously developed by Chang, et al. (*14*) for the simulation of vehicular movements on highways and arterials. The simulation logic follows that of the magneto-hydrodynamic code developed by Tajima et al. (*15*) for plasma physics applications. It is a fixed time-step simulator. Vehicles on a link are moved individually at prevailing local speeds consistent with macroscopic speed-density relations (modified Greenshield's model in this case). Inter-link transfers are subject to capacity considerations. For given network representation and link characteristics, the simulator takes a time-dependent input function and determines the associated vehicular movements, thereby yielding the resulting link trip times, including estimated delays associated with queueing at nodes. These form the input to the path search and assignment component, which calculates the pertinent path trip times, which are in turn used by the user decisions component. The latter is intended to predict the response of users to the available information, according to a set of behavioral rules of the kind described in the previous section. Another function of the second component is to translate the user path selection and switching decisions into time-varying link flow patterns on the network's links. Further detail on the simulation-assignment methodology can be found in the paper by Mahmassani and Jayakrishnan (*2*).

All simulation experiments were performed on a CRAY X-MP supercomputer to meet the extensive memory requirements associated with applying the behavioral rules at the individual vehicle level and tracking individual vehicle paths.

Experimental Factors

The simulation experiments were conducted to explore the opportunities for system improvement with respect to four principal factors—information source, behavioral rule, loading pattern, and market penetration. These are described in turn hereafter.

Information Source

As mentioned previously, two information sources are considered: home-based information, consulted before actually starting the trip, and in-vehicle en route information. Four strategies are considered for this factor: (*a*) no information (base case), (*b*) home-based pretrip information only, (*c*) en route only, and (*d*) both sources are available. Under the second strategy (*b*), path selection is allowed only at the origin. Once users are driving on any given link (including the entry ramp), it is no longer possible for them to switch to another facility (at least not in response to real-time information in this model). Under the third strategy (*c*), users only have access to information along the way. They always enter the corridor system through their individual primary highway path (assigned as part of the loading pattern) and can switch through crossover links only. Under the last strategy, users have access to real-time information both at the origin and en route, and can therefore select their initial path accordingly as well as switch paths along the way. As explained previously, the relative effectiveness of these two sources of information is one of the principal objectives of this paper.

Behavioral Rule

Only the results for the myopic rule (R.1) and the bounded-rational relative indifference band (R.2) are presented in this paper. In the simulation runs, each user with information is assigned a randomly generated indifference band, η_i, drawn from a triangular distribution with mean $\bar{\eta}$ and range $\bar{\eta}/2$. In Mahmassani and Jayakrishnan's previous simulation experiments (*2*), it was found that a mean indifference band of 0.2 appears to provide reasonable overall behavior as well as the largest systemwide improvement in travel time. Thus only two levels of $\bar{\eta}$ were considered in these experiments: 0.0 and 0.2. In the no band case ($\bar{\eta} = 0.0$), all users are assumed to have a zero band, hence this case represents the myopic case (rule R.1)—users with information will always switch to an alternative path if it offers an improvement in trip time, no matter how small. The minimum absolute improvement threshold τ_i, set at 1 min, is taken to be identical across all users with information, except in the zero band case in which no minimum improvement restriction is imposed.

Two different mean indifference bands were implemented in these experiments: $\bar{\eta}_{1i}$ for the pretrip route choice model and $\bar{\eta}_{2i}$ for the en route path switching model. Hereafter, denote each case by $[\bar{\eta}_{1i}, \bar{\eta}_{2i}]$ where a value of 99 is used to indicate that no switching is allowed. (In the actual computer program, the minimum absolute improvement threshold τ_i was set at a very high value to preclude any switching.) Thus two cases each were considered for home-based pretrip switching only and en route switching only. These are [0.0, 99] and [0.2, 0.9] for the former, and [99, 0.0] and [99, 0.2] for the latter. Four combinations of behavioral rules were considered for the situation where both home-based and en route information are available: [0.0, 0.0], [0.0, 0.2], [0.2, 0.2], and [0.2, 0.0]. Of course, the no-information base case corresponds to [99, 99].

Loading Pattern

One of the principal determinants of the existence of opportunities to improve the overall system consists of the existing network traffic conditions. To capture this effect, three loading patterns were used in the simulation experiments. These are conveniently referred to hereafter as loading patterns 1, 2, and 3. In all cases, a total of 10,800 commuters, split equally among the first six (residential) sectors, share the use of facilities in the corridor during the morning commute. Commuters in each sector depart uniformly over a 20-min period; the loading periods for each sector are staggered with a time lag of 5 min between adjacent sectors, with Sector 1 starting first.

Under the first loading pattern, commuters are split equally among the three highways, departing at a rate of 30 veh/min/sector for each facility. The second loading pattern has departing rates of 40 veh/min for Highway 1, 30 veh/min for Highway 2, and 20 veh/min for Highway 3, for each sector. Under loading pattern 3, 60 vehicles enter Highway 1, 20 vehicles enter Highway 2 and 10 vehicles enter Highway 3, all min/sector. Note that these assignments constitute intended paths for the commuters. If origin-based real-time information is available, the actual initial path selected by commuters with access to such information may be different.

Market Penetration

To examine the effect of this critical parameter, five levels of the fraction of users with access to real-time information were considered, spanning the spectrum from luxury gadget to universal availability: 0.10, 0.25, 0.50, 0.75, and 1.00. As they are generated, individual vehicles are assigned their information availability status randomly and independently according to the above fractions.

Using different combinations of these four experimental factors, 123 separate simulation runs were performed. The results are discussed in the next section.

ANALYSIS OF RESULTS

System performance for each simulation run is evaluated by comparing the average trip time for all commuters in the system to the corresponding value in the base case (the do-nothing case, denoted by [99, 99] in the previous section, with the same initial loading pattern). The average trip times for the base case for loading patterns 1, 2, and 3 are 23.30 min, 21.45 min, and 23.26 minutes, respectively. The clearly superior performance of the system under loading pattern 2 reflects the assignment of vehicles in relative proportion to the individual facilities performance characteristics, as captured here by the free mean speed (recall that Highway 1 has the highest free mean speed). Loading pattern 2 thus appears to provide an assignment of vehicles that is closer to some optimal value than the other two. On the other hand, loading pattern 1 has relatively too many vehicles on Highway 3, whereas loading pattern 3 underuses Highways 2 and 3 while overloading Highway 1.

Figure 2 depicts the variation in total trip time in the system (expressed as a percent of the base case) with the fraction of the user population with access to information, under both myopic and boundedly rational indifference band rules, for each of the three loading patterns, under real-time home-based information availability only. Each case is identified in the figure legend by the notation defined in the previous section. Figure 3 presents similar results for the situation where only en route real-time information is available, whereas Figure 4 presents the results of the simulations under both home-based and en route information availability. Because the total trip times are expressed as a percent of the base case under that particular loading pattern, values exceeding 100 percent indicate a worsening of systemwide performance (compared with the do-nothing case).

To capture the incidence of the benefits (and/or costs) on the user population, Figures 5, 6, and 7 depict the average trip time experienced by those who have access to information in the three cases considered, respectively, again expressed as a percentage of the corresponding average trip time in the do-nothing base case. Figures 8, 9, and 10 present similar information for those trip makers with no access to information. The system's overall performance is examined next, followed by a discussion of the incidence of the benefits and costs on those with and without access to information.

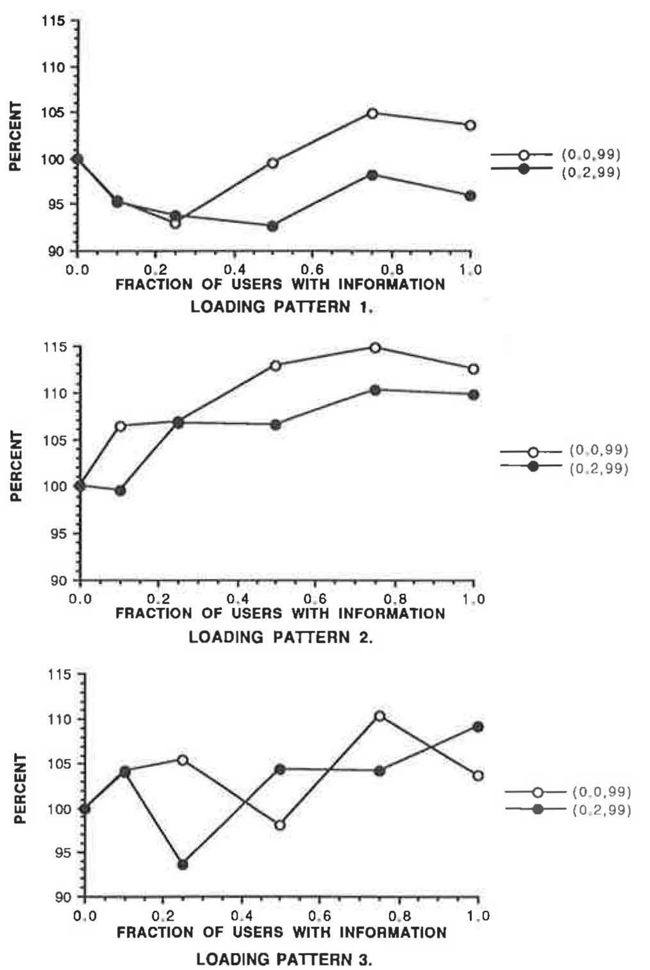


FIGURE 2 Variation of average trip time for all users, as a percentage of no-information base case, under home-based real-time information availability only.

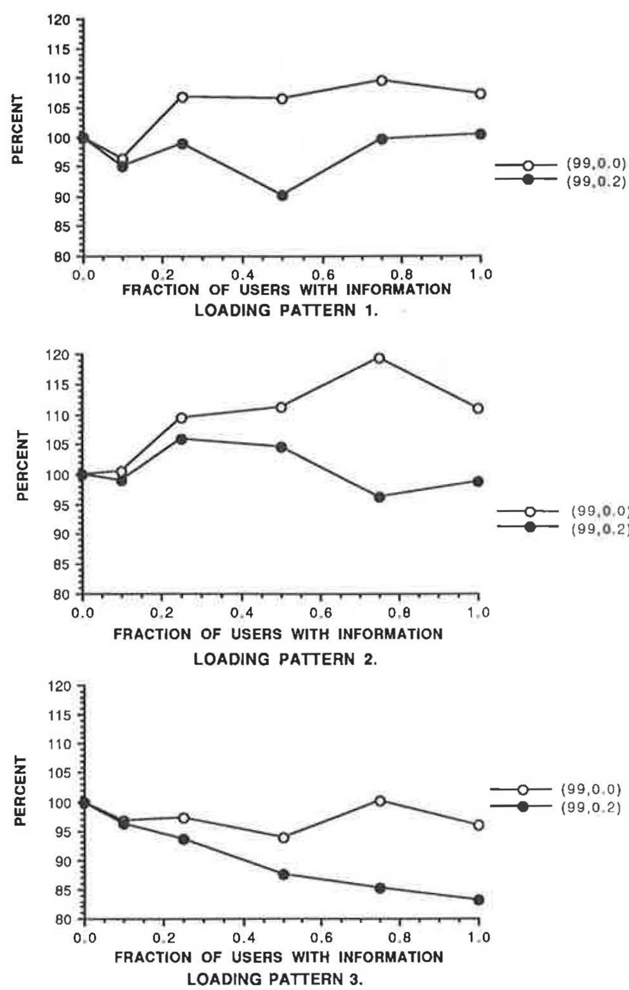


FIGURE 3 Variation of average trip time for all users as a percentage of no-information base case, under en route real-time information availability only.

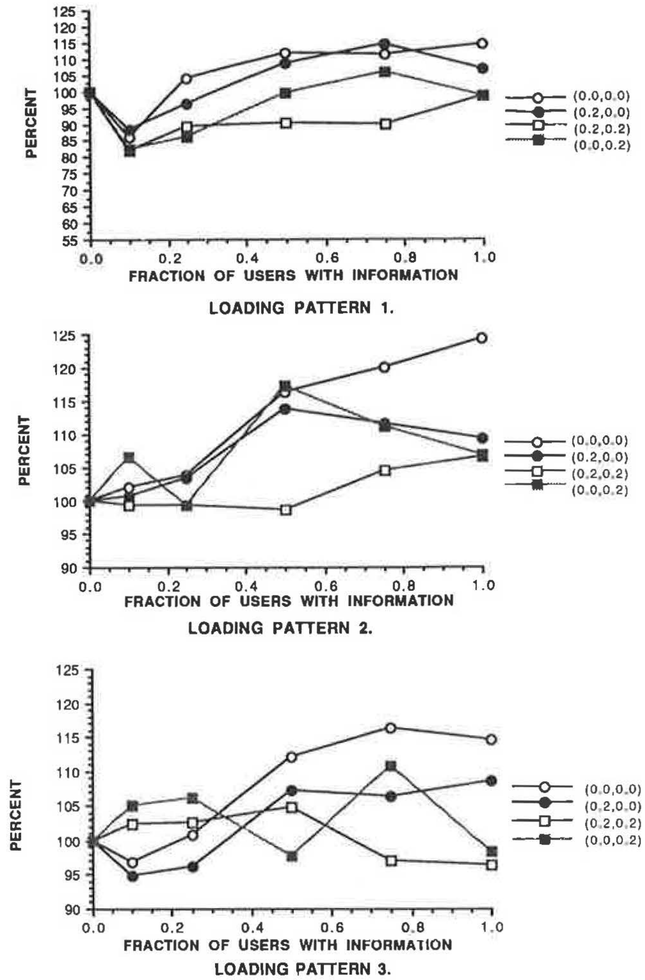


FIGURE 4 Variation of average trip time for all users, as a percentage of no-information base case, under both home-based and en route information.

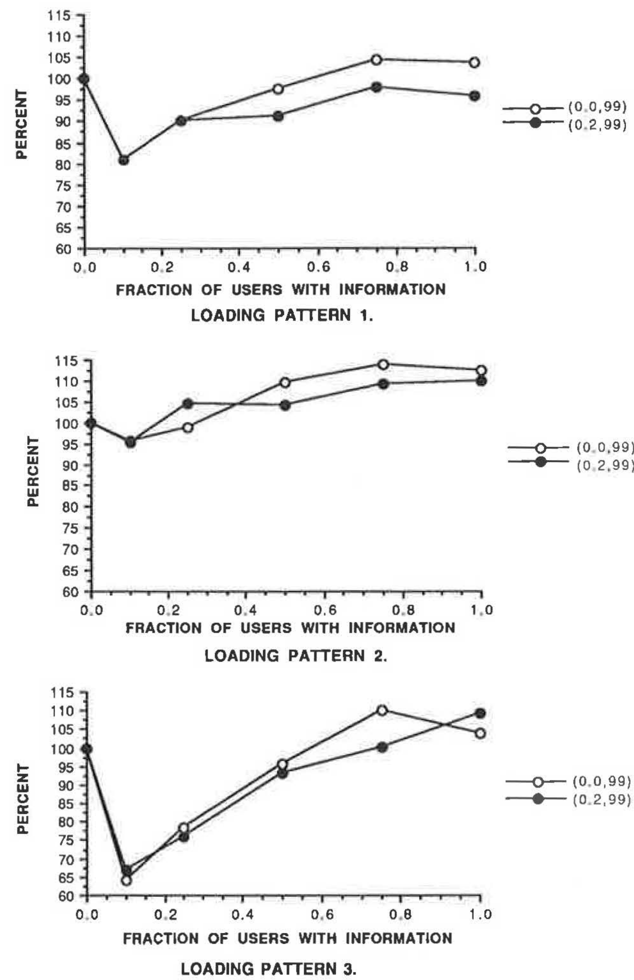


FIGURE 5 Variation of average trip time for users with information, as a percentage of no-information base case, under home-based real-time information availability only.

Systemwide Performance

From a system controller's standpoint, the purpose of home-based real-time information would be to rearrange the loading pattern, through pretrip path selection, in such a way that it becomes closer to the system optimum. However, this information source may not necessarily succeed in improving the system, especially when users select their respective paths according to the myopic rule (R.1), as illustrated in Figure 2. Improvement in system performance is observed for up to 50 percent market penetration, with a high of about 7.0 percent improvement, attained at 25 percent market penetration, and a low of 4.9 percent increase in delay, at 75 percent market penetration under the first loading pattern. Virtually no improvement is obtained under loading pattern 2, with system performance actually worsening by a maximum of approximately 15 percent at 75 percent market penetration. Very little improvement in overall performance is attained when users respond according to the myopic rule (R.1) under initial loading pattern 3. The highest improvement achieved under any market penetration level considered is only 2.0 percent, at 50 percent market penetration. For all other levels, the performance worsens under rule R.1, to a maximum of approximately 10.4 percent.

Under the boundedly rational response rule (R.2), with a mean indifference band of 0.2, Figure 2 indicates that, for loading pattern 1, system performance under origin-based real-time information improves at all levels of market penetration, with a maximum of approximately 7.4 percent. Since loading pattern 2 is closer to the optimum, there appears to be little room for improvement through pretrip path assignment. In fact, although the average trip times get better across the board relative to the myopic cases, they are still worse than the no-switching base case, except at very low levels of market penetration in which only marginal overall improvement is attained. On the other hand, the indifference band buffer under R.2 does not necessarily lead to better overall performance relative to the myopic rule (R.1) for loading pattern 3. A notable instance where it does is at the 25 percent market penetration level, with an overall improvement of 6.3 percent.

Comparing the results in Figure 2 with those in Figure 3, it appears that the system for loading pattern 1 performs somewhat better under origin-based information only than

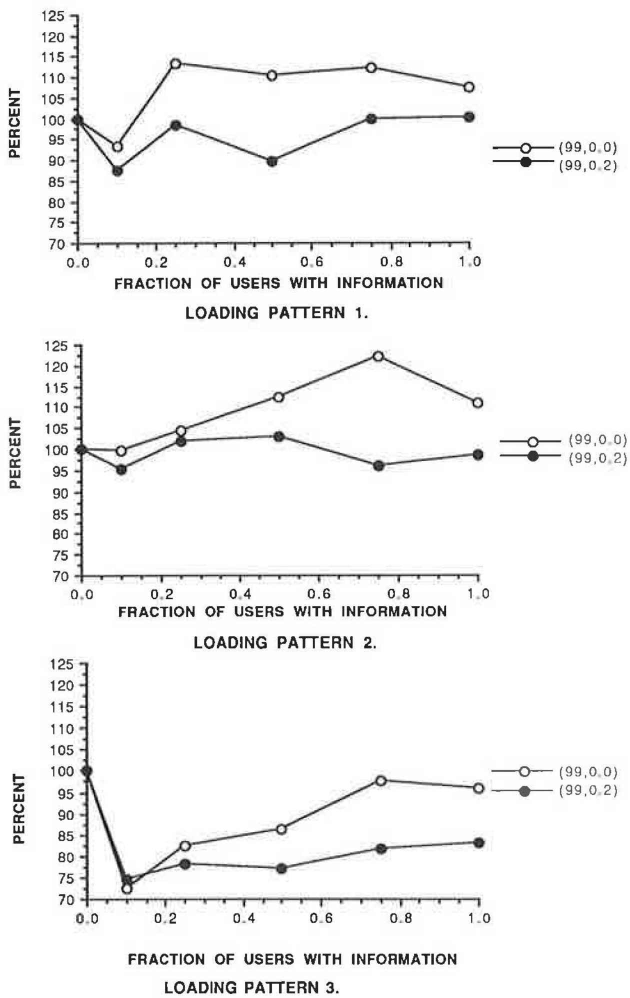


FIGURE 6 Variation of average trip time for users with information as a percentage of no-information base case, under en route real-time information availability only.

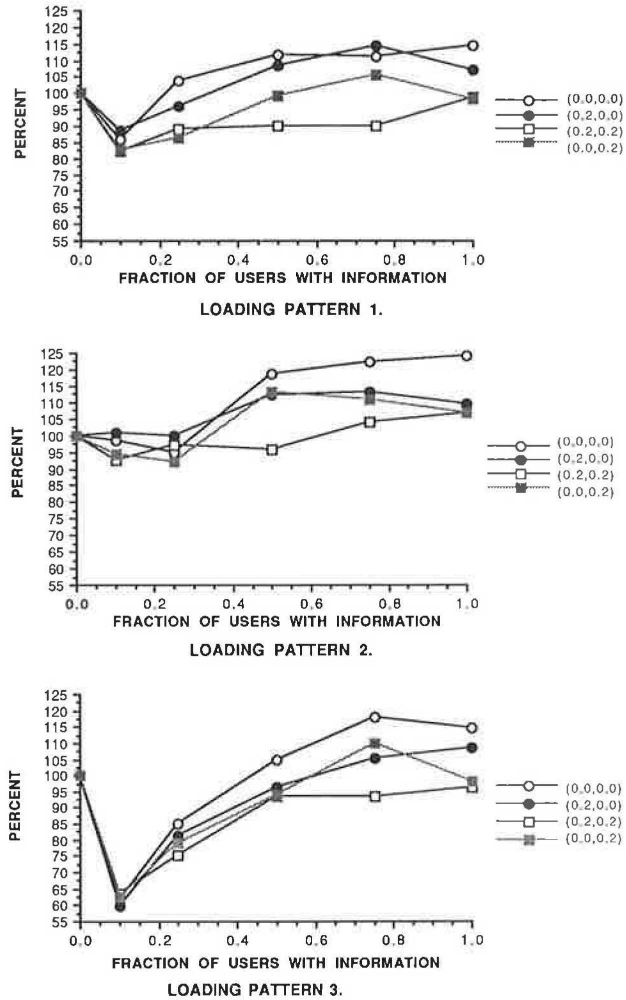


FIGURE 7 Variation of average trip time for users with information, as a percentage of no-information base case, under both home-based and en route information.

under en route information only. This is true primarily under the myopic rule (R.1), whereas R.2 yields equivalent or slightly better performance under en route switching at lower market penetration levels, but somewhat higher trip times when most users have access to information. On the other hand, the system performs clearly better under en route than under origin-based information for loading pattern 3. One possible explanation has to do with the direction of the changes necessary to improve each of the two loading patterns, which makes them more or less robust vis a vis incorrect or poor switching decisions. For instance, loading pattern 3 is more susceptible to worsening under pretrip path selection only because its improvement would require redistribution of flow from the fastest facility (at free flow), Highway 1, to the other two. Because in the early stages of loading, Highway 1 will still outperform the other two, it will receive diverted traffic from the other two facilities under origin-based information. This diverted traffic will increase congestion downstream on Highway 1, especially as downstream departures start at their already heavy rate under pattern 3. Under origin-based path selection only, no opportunities exist to redistribute flows downstream, precluding the ability to reverse the negative impact of earlier poor switches. On the other hand, to improve loading pattern 1, flow needs to be redirected from the slowest to the fastest facility (at free flow). Therefore, the initial diversion from Highway 3 to Highway 1 under origin-based information would be in the correct direction. Of course, under high levels of market penetration, collective effects lead to reduced effectiveness and possible worsening, even under home-based information only.

Figure 3 also reveals that overall system performance is clearly better when users dampen their response to en route real-time information through a minimum threshold, as in Rule R.2, than when they respond according to the extreme myopic rule R.1. This is true for all three initial conditions considered, but especially for loading pattern 3, for which the improvement continues to increase, albeit at a decreasing rate, with the level of market penetration, up to a systemwide improvement of 16.8 percent over the do-nothing base case. Furthermore, less extreme behavior by users, such as under R.2, tends to dampen the negative impact of the collective effects that appear at high levels of market penetration.

Conceptually, providing users with both origin-based and en route information would be similar to attempting to modify

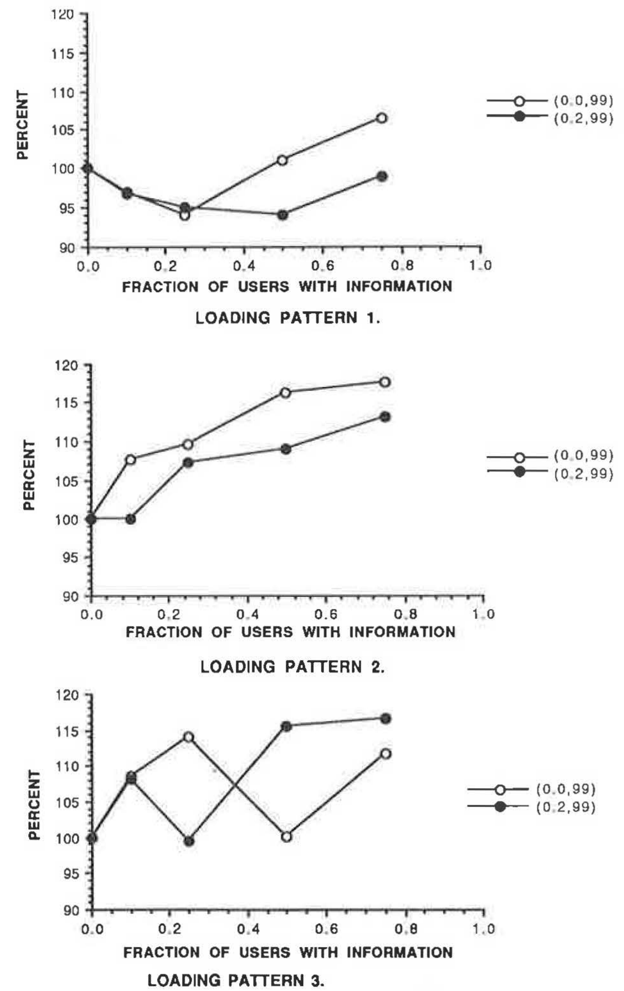


FIGURE 8 Variation of average trip time for users without information as a percentage of no-information base case, under home-based real-time information availability only.

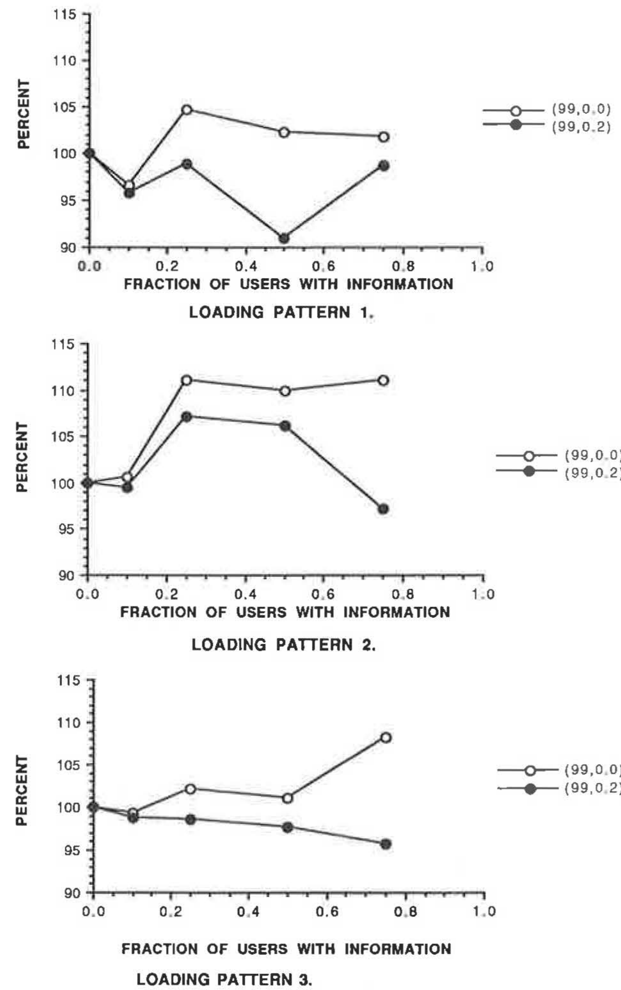


FIGURE 9 Variation of average trip time for users without information as a percentage of no-information base case, under en route real-time information availability only.

the initial conditions (loading pattern) through pretrip switching of users with information, and then seeking to improve the system further by en route switches. However, the analysis of the results presented for en route switching only (Figure 3) have indicated that systems with initial conditions close to the optimum leave little room for improvement, as in the case of loading pattern 2, and may actually get worse, especially at high market penetration levels. Therefore, providing both sources of information appears to reduce the potential effectiveness of either source taken individually, as shown in Figure 4. In many cases, for all three loading patterns, especially under extreme user behavior (rule R.1), the potential negative effects of the two sources appear to be compounded. The worst case is attained at 100 percent market penetration under loading pattern 2 using the myopic rule for both initial and en route switching, with approximately 24.2 percent increase in systemwide trip time. In general, better systemwide system performance is attained when path switching behavior is dampened by an indifference band, both initially as well as en route (i.e, the [0.2, 0.2] case).

Incidence of Benefits and Costs

One of the fundamental issues that transportation systems engineers and policy makers have to deal with is the incidence of the benefits and costs of information on users with and without access to such information. The associated equity issues may well be among the toughest obstacles facing the implementation of advanced driver information systems. With this in mind, average trip times have been computed separately for users with and without information for all the cases considered. The results are summarized in Figures 5, 6, and 7 for those with information and in Figures 8, 9, and 10 for the other group, as noted previously.

Figures 5, 6, and 7 suggest that average users with access to information (from either source) will reduce their trip time in all cases considered at low market penetration levels. The relative advantage gained by users with information appears to be highest when access to such information is limited to about 10 percent of the commuter population. This improvement in the trip time of those with information was achieved

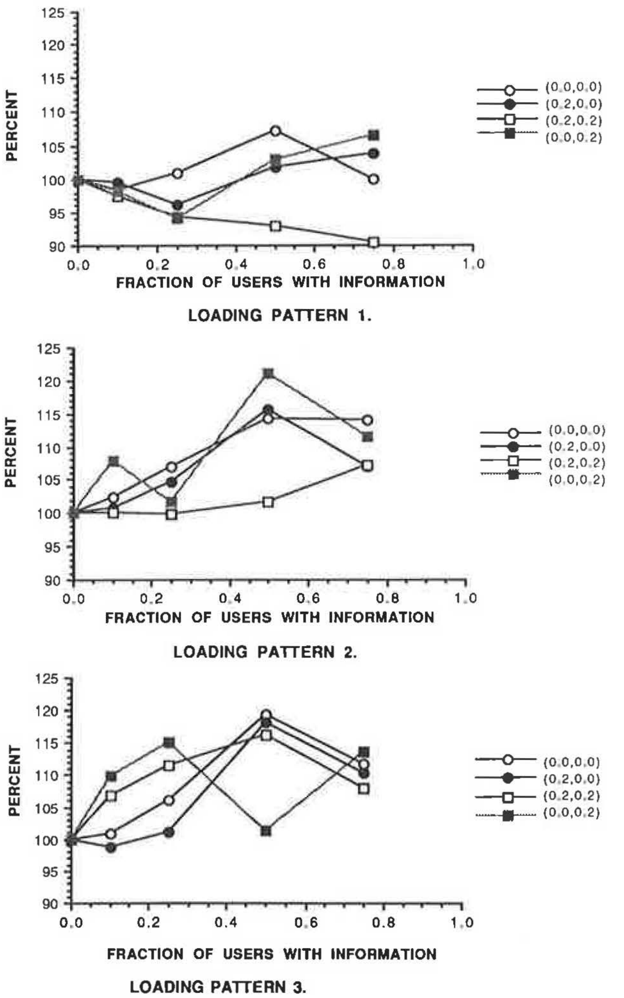


FIGURE 10 Variation of average trip time for users without information as a percentage of no-information base case, under both home-based and en route information.

at low market penetration levels even when systemwide performance actually worsened. However, as more users have access to information, their trip time savings decrease. In some cases, especially under loading pattern 2, users with information actually experience longer trip times than they would have under the no-information base case. Furthermore, it appears that the closer the initial system is to the system optimum (e.g., loading pattern 2), the lower is the market penetration level beyond which even users with information actually do worse than under no-information (this level is about 15 to 20 percent for loading pattern 2, under either source of information).

Under departure pattern 3, whereas no meaningful systemwide benefits were achieved with home-based information only, the informational advantage resulted in very significant savings at low penetration levels. The most improvement of any case was achieved by those with such information under this loading pattern (40.1 percent reduction in trip time at 10 percent penetration). However, these benefits diminish dramatically as market penetration increases, resulting in an eventual worsening at market saturation. Under loading pattern 1, users with information experience less benefit than under pattern 3 at the 10 percent penetration level. However, this benefit does not decrease as fast with increasing market penetration.

In most cases, home-based information only yields greater savings to those with access to it than en route information only at low market penetration levels (not to exceed 15 to 20 percent for the scenarios simulated in this paper). However, the benefits of the home-based only information source rapidly become diluted as market penetration increases, and eventually become nonexistent or negative as one approaches market saturation. On the other hand, the benefits of the en route only source (to those who have access to it) appear to be somewhat more robust with respect to increasing market penetration. Nevertheless, beyond some level that is highly dependent on initial conditions, these benefits may disappear or turn negative.

When both sources of information are used by commuters, additional benefits, relative to those achievable under either source separately, accrue to users with information at low market penetration levels. However, the benefits are far from being additive. The incremental benefits are rather limited to a few percentage points (relative to the base case). However, these benefits rapidly decrease with market penetration, and are often less than those attainable under one source only.

The results of Figures 8, 9, and 10 for users without information reveal that only in a few cases does this group actually experience a reduction in average trip time (relative to the no-information base case). Most of these cases occur under loading pattern 1. Moreover, in several cases, this group does somewhat better at higher market penetration levels, as the larger number of those with information divert to the supposedly better paths, leaving mostly the uninformed drivers to experience the benefits of reduced congestion on the relieved facilities. However, this does not occur in all cases, and appears to be dependent on initial conditions in the system. Furthermore, benefits which may accrue to the uninformed drivers are typically quite small.

CONCLUSIONS

The simulation experiments presented in this paper are limited to a particular network configuration under three different initial conditions (loading patterns), as well as to particular real-time information supply sources and strategies. As such, considerable caution must be exercised in interpreting the results and before attempting to generalize their applicability. Nevertheless, important and useful insights have been obtained into the effects of real-time information advisory sources on overall traffic system performance, some of the critical factors that influence this performance, and the dependence of these effects on the behavioral mechanisms governing the response of users to the supplied information.

It is important to recall the nature of the information supplied to users. Descriptive real-time information on currently prevailing link trip times coupled with a capability to compute self-optimized paths were considered. Under home-based information, this capability is available at the origin of the trip, before the driver actually gets on the network. With en route information, trip times for paths from the driver's current node to his or her destination are updated on a quasi real-time basis. It is important to keep in mind in interpreting the results of the simulations, the trip times used in the calculation that form the basis of the user decisions are those that are currently prevailing in the network. Of course, these may or may not be the travel times actually experienced by the users when they traverse these links. In other words, no attempt is made to predict what the travel time will be on a given link in the future, taking into account the path choices of the users in response to the supplied information.

It was strongly suggested in the results that the overall effectiveness of real-time information in a network is highly dependent on the prevailing initial conditions, and the extent to which these offer opportunities for improvement. It appears that the closer the system is to the optimum, the higher the likelihood that information from either source may actually worsen systemwide performance. Under loading pattern 2, the best average systemwide trip time is 20.66 min, and the worst is 26.65 min, as opposed to that for the no-switching base case, which was 21.45 min. On the other hand, meaningful improvement was obtained when the initial conditions were further away from the optimum. For instance, under loading pattern 3, the best system performance attained was 19.36 min, which outperforms the best case under loading pattern 2.

The relative effectiveness of home-based versus en route information is also highly dependent on initial conditions, and on the manner in which the present system may be suboptimal. For instance, it was seen that when the fastest facilities (e.g., freeways) are underutilized, origin-based information could be effective. When the reverse is true, initial switching appears to be effective only at very low market penetration levels, and en route information seems to be much more effective. Actual systems are likely to be more like loading pattern 3—to overutilize faster facilities—suggesting that en route information would be preferable from a systemwide standpoint. Of course, this conclusion does not take into account the longer term adjustments that might take place through day-to-day changes in route and departure time. Apparently, if present conditions are already close to the optimum, then the indiscriminate provision of information, both origin-based as well as en route, may actually worsen conditions through users' overreaction and myopic shifts.

Again, it is important to emphasize the limitations arising from the particular network configuration considered, and the manner in which the information was provided to the users in our simulations. However, the results do suggest the need to carefully consider several important parameters and factors in the ongoing research and development efforts pertaining to advanced driver information systems. In particular, the following four items should be highlighted because of the results of our experiments, as well as those that are beginning to emerge in related research.

1. The importance of initial conditions in determining the potential effectiveness of real-time information strategies has been highlighted in previous research by Mahmassani and Jayakrishnan (*2*), and Mahmassani, et al. (*16*). The present results further confirm this conclusion and offer insights into how the character of the initial conditions affects the impacts of information strategies from different sources. Additional effort should be directed at characterizing present conditions in congested networks, especially in terms of how tripmakers actually utilize the components and facilities of these networks.

2. The potential negative effects of extreme behavior in the users' response to the supplied information are further documented in our results. The existence of benefits from advanced driver information systems obviously depends on the manner in which users respond to the information. Allowing for bands or thresholds in one's decision to switch to a presently better alternative clearly improves systemwide performance as well as the benefits that might accrue to those with the informational advantage. Although this matter is not under the system controller's direct control, it can be influenced through the design of the supplied information itself as well as though possible education of the driving public. Ultimately, it is likely that users themselves will reach their own conclusions about appropriate switching rules, through repeated experience with the facility. The dynamics of the formation of such indifference bands constitute an important subject of additional research, which could benefit from previous work on the day-to-day dynamics of commuting decisions through the use of laboratory experiments (*7*).

3. The need for coordination in the provision of information beyond a certain market penetration level is strongly suggested by our results. This point has been recognized in proposals for ADIS implementation programs (*1*). Our results suggest that the level beyond which coordination is needed may be as low as 10 or 20 percent, depending on initial conditions.

4. The nature of the information supplied to the motorists should receive considerable attention. As emphasized throughout, our results are predicated on current trip times and no attempt is made at prediction. However, the strong interrelation among supplied times, user decisions, and traffic conditions makes the prediction problem and the design of information supply strategies rather complex. This item also ties in with the preceding one regarding the development of coordinated information and control strategies.

ACKNOWLEDGMENTS

Funding for the work on which this paper is based comes from a grant from the Advanced Technology Program of the Texas Higher Education Coordinating Board, and from the U.S. Department of Transportation through the Southwest Region University Transportation Center. The work was initiated with funding from the General Motors Research Laboratories. The contributions of R. Jayakrishnan to several aspects of this work are gratefully acknowledged. This work has also benefited from fruitful interaction with Srinivas Peeta, Robert Herman, and Richard Rothery. The views expressed herein are solely those of the authors and do not necessarily reflect those of the sponsoring agencies.

REFERENCES

1. Mobility 2000 (1990) "Advanced Driver Information Systems (ADIS)," Final Report of the Working Group on ADIS.

2. H. S. Mahmassani and R. Jayakrishnan. "System Performance and User Response Under Real-time Information in a Congested Traffic Corridor," forthcoming in *Transportation Research*, Vol. 25, No. 5, 1991.
3. Y. Sheffi. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Englewood Cliffs, N.J., 1985.
4. H. S. Mahmassani and G. L. Chang. Dynamic Aspects of Departure Time Choice Behavior in a Commuting System: Theoretical Framework and Experimental Analysis. In *Transportation Research Record 1037*, 1985, pp. 88–101.
5. H. S. Mahmassani and G. L. Chang. On Boundedly-Rational User Equilibrium in Transportation Systems. *Transportation Science*, Vol. 21, No. 2, 1987.
6. H. S. Mahmassani and D. G. Stephan. Experimental Investigation of Route and Departure Time Choice Dynamics of Urban Commuters. In *Transportation Research Record 1203*, 1988, pp. 69–84.
7. H. S. Mahmassani and R. Herman. Interactive Experiments for the Study of Tripmaker Behaviour Dynamics in Congested Commuting Systems. In *Developments in Dynamic and Activity-Based Approaches to Travel Analysis* (P. Jones, ed.), Gower, Aldershot, England, 1990, pp. 272–298.
8. C. F. Daganzo and Y. Sheffi. On Stochastic Models of Traffic Assignment. *Transportation Science*, Vol. 11, No. 3, 1977.
9. Mobility 2000. *Operational Benefits*. Final Report of the Working Group on Operational Benefits.
10. L. Hendrickson and E. Plank. The Flexibility of Departure Times for Work Trips. *Transportation Research A*, Vol. 18A, 1984, pp. 25–36.
11. H. S. Mahmassani and G. L. Chang. Experiments with Departure Time Dynamics of Urban Commuters. *Transportation Research B*, Vol. 20B, No. 4, 1986, pp. 297–320.
12. G. L. Chang and H. S. Mahmassani. Travel Time Prediction and Departure Time Adjustment Behavior Dynamics in a Congested Traffic System. *Transportation Research B*, Vol. 22B, No. 3, 1988, pp. 217–232.
13. C. C. Tong, H. S. Mahmassani, and G. L. Chang. Travel Time Prediction and Information Availability in Commuter Behavior Dynamics. In *Transportation Research Record 1138*, 1987, pp. 1–7.
14. G. L. Chang, H. S. Mahmassani, and R. Herman. A Macroparticle Traffic Simulation Model to Investigate Peak-Perid Commuter Decision Dynamics. In *Transportation Research Record 1005*, 1985, pp. 107–120.
15. J. Leboeuf, T. Tajima, and J. Dawson. A Magnetohydrodynamic Particle Code for Fluid Simulation of Plasmas. *Journal of Computational Physics*, Vol. 31, No. 3, 1979, pp. 279–408.
16. H. S. Mahmassani, R. Herman, C. M. Walton, E. G. Jones, M. H. Baaj, and R. Jayakrishnan. *Travel Time Characteristics and Opportunities for Real-Time Information Systems in an Urban Commuting Corridor*. Report GM-1989-F, Center for Transportation Research, The University of Texas at Austin Tex., 1989.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Laboratory Assessment of Driver Route Diversion in Response to In-Vehicle Navigation and Motorist Information Systems

R. Wade Allen, David Ziedman, Theodore J. Rosenthal, Anthony C. Stein, Jaime F. Torres, and Abolhassan Halati

A laboratory investigation of driver use of in-vehicle navigation systems is described in this paper. This study is the first phase of a two part project in which the second phase will apply the driver behavior data to a traffic simulation model. The objective of the driver behavior experiment was to compare the effect of four navigation systems on driver diversion decisions when faced with traffic congestion. Three of the systems were developed on the basis of a heading-up map display. These systems varied from a basic map with vehicle position information to a highly complex map with position, congestion, and route guidance information. The fourth system consisted of simplified symbolic directions and distance to change information. The experiment simulated typical freeway trips using sequences of slides of real freeway scenes and auditory feedback controlled by a computer. Drivers were presented information on traffic congestion, vehicle speed and guide signs of off-ramps, and were motivated with monetary rewards and penalties to encourage diversion decisions that would minimize trip travel delays. In addition to several in-vehicle navigation system configurations, experimental variables included driver route familiarity, age group, and either commercial or noncommercial driving experience. The results showed that navigation system characteristics can have a significant effect on driver diversion behavior, with better systems allowing more anticipation of traffic congestion. This result was found over several different levels of congestion. Driver age also was a factor, with old drivers more reluctant to divert from the main freeway route. Route familiarity, commercial driving experience, and gender group variables were not significant factors in driver diversion decision making.

Traffic congestion in the United States causes millions of dollars in lost revenue and millions of hours in lost time every year. In addition, King (*1*) reports that driver navigational waste is equal to 6.4 percent of all distance and 12.0 percent of all time spent in travel by noncommercial motorists. With the increasing number of vehicles on the road, the movement of workers from the city, and the highway network in the United States nearly completed, new solutions to traffic congestion are necessary. The application of available technology, including appropriate human factors design, could potentially provide a large part of the solution to traffic congestion. Driver behavior in response to one potential technical solution, the use of in-vehicle navigation systems, is examined in this paper. A purpose of these systems is to allow drivers to avoid traffic congestion. The main point addressed in this paper is how drivers respond to added navigation information when confronted with traffic congestion.

Various classes of navigation systems have been defined from Class 0 open-loop systems, which are basically autonomous, to Class 4 dynamic closed-loop systems (*2*). Open-loop systems include simple directional aids, map display systems, and route guidance aids. The Etak Navigator is a currently available autonomous electronic map-based system which includes vehicle and destination position information (*3*). A dynamic closed-loop system contains two-way communication between vehicle and control center (Class 4). Here, centralized vehicle tracking, optimal routing, and information transfer to the vehicle are included such as with the Ali-Scout system (*4*). A system with moderate sophistication (Class 3) might include one-way communication with traffic congestion information being transmitted to the driver.

In developing the technology for navigation systems, several important issues arise associated with system performance and safety of operation. The focus of the human factors study described herein is on driver behavior associated with in-vehicle navigation systems from a system performance point of view. Will in-vehicle navigation systems encourage drivers to take alternate routes and divert early to avoid congestion? The study documented herein approached the testing by simulating trips using slide representations of the freeway environment and prototypes of various navigation system formats. It was hoped that a range of applicable results could be obtained which would otherwise be prohibitive in cost using more expensive simulation techniques or on-the-road evaluations. Human factors principles were used to define the prototype navigation displays, but display design and driver-display interface issues were not explored in detail in this research.

BACKGROUND

Psychological and human factors guidelines for the design of in-vehicle navigation systems have not been established. However, a literature review (*5*) uncovered several issues that should be taken into account. A driver's ability to navigate

R. W. Allen, D. Ziedman, T. J. Rosenthal, and A. C. Stein, Systems Technology, Inc., Hawthorne, Calif. 90250-7083. J. F. Torres and A. Halati, JFT Associates, Pacific Palisades, Calif. 90272-1207.

through a complex environment is largely dependent on the type and extent of cognitive structures representing that environment, the goals of the driver, and the ability of the driver to stay oriented. These three areas, founded in psychology and environmental cognition, are functionally related. First, a destination and travel plan must be formed. Second, knowledge of the local or global network must be known or acquired. Finally, a reference system must exist to relate the driver to the environment.

The cognitive map has been hypothesized as the basis for mentally storing or representing information about the physical world we navigate in (*6*). The internal format for remembering this information could have profound effects on the ease with which one can assimilate information presented by a navigation system. If the information is mentally stored in a prepositional format, then specific verbal directions may be desirable (*7*). However, if the information is in a format analogous to the real world, a different representation, the map for example, may be desired (*8*). In addition, the spatial and verbal skills of drivers may vary significantly among individuals, thereby influencing their ability to use different navigation display formats (e.g., *10*).

Human factors issues of concern include the format and coding of navigation system information (*11*), the attentional demand and safety issues of displays and controls (*12*), and agreement on general guidelines for the development and manufacture of navigation systems. Research into these issues is in its infancy. These issues were considered here in terms of laying out prototype navigation system displays, but were not addressed otherwise in the research. Another important navigation issue is the route choice behavior of drivers. When travel planning before or during a trip, several key variables influence driver decisions. These include cost, total trip time, delay time, distance, trip purpose, and traffic congestion levels (*1*). Many of these variables may be based on secondary environmentally related variables including route complexity, perceived average speeds, number of traffic signals or stop signs, number of lanes, and so on. The weights given to these variables affect a driver's route choice and hence they should be taken into account in designing any navigation system.

The design of the prototype navigation systems and testing procedures reported on herein took these issues and variables into account. Although the prototype display designs are probably not optimal, an attempt was made to follow good human factors interface design practice within the limitations of the PC computer system available for the research. The objective was to achieve prototype systems containing the necessary information to adequately test content and design differences.

APPROACH

Driver use of prototype in-vehicle navigation systems was measured with a part task simulation. The simulation, as described in the following, presented subjects with several traffic congestion scenarios in which they attempted to avoid congestion delays using prototype navigation system information. Four subject groups used different navigation system configurations and a control group was not given any navigation information. The laboratory simulation and experimental procedures were designed to motivate subjects to avoid heavy congestion as they would in the real world.

Navigation System

Four prototype navigation systems were defined for testing that gave varying amounts of information on alternate routing and congestion conditions. Example display conditions are shown in Figure 1 and the basic system capabilities were as follows:

- Static map system—map display with vehicle position indicated (no congestion information was provided by this system);
- Dynamic map system—a static map with traffic congestion level information;
- Advanced experimental system—a dynamic map with highlighted alternate route, additional textual information, and auditory instructions; and
- Route guidance system—a non-map-based system using arrow symbols for direction instructions, a bar graph representing distance to exit, estimated arrival time, and distance to destination.

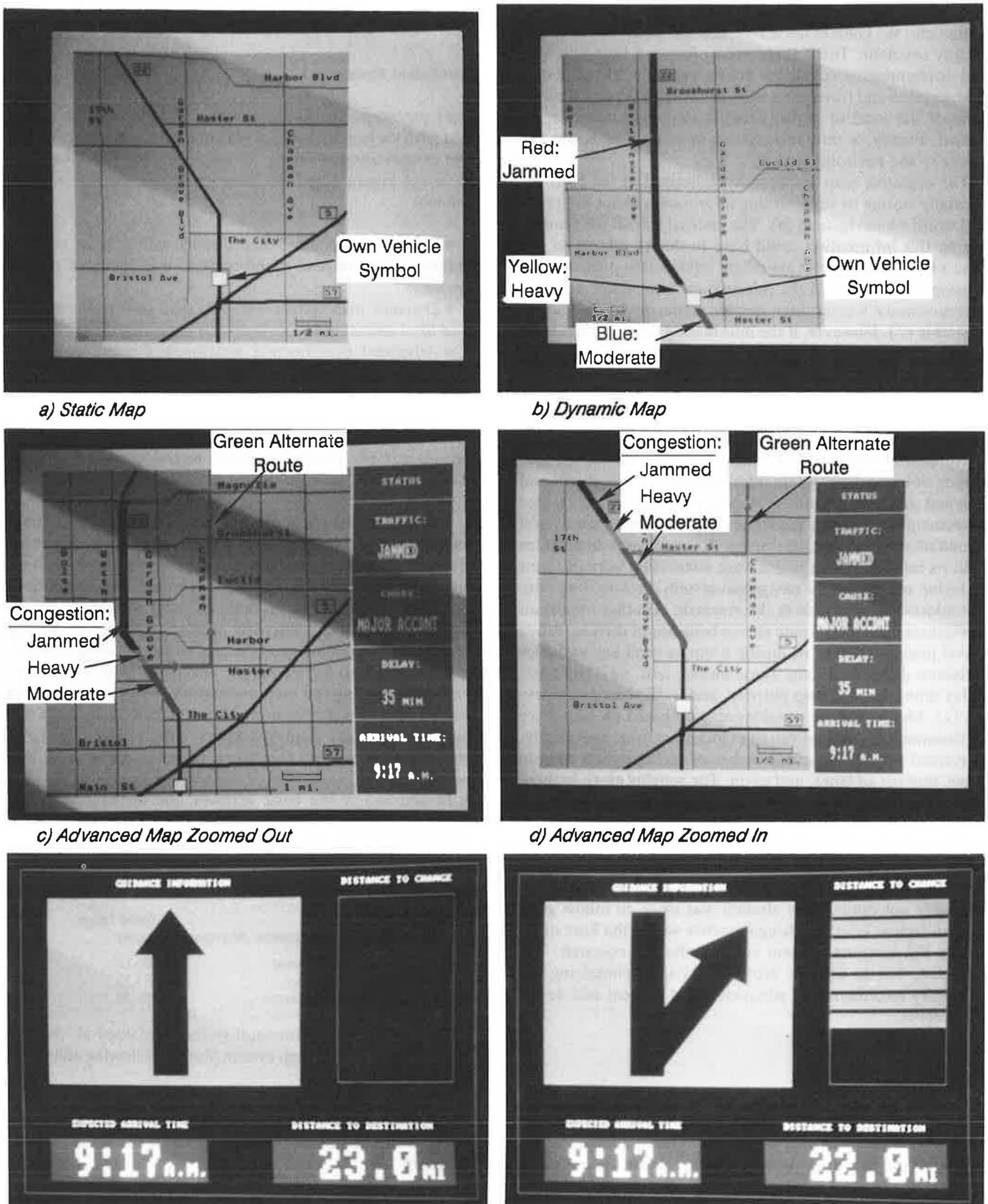
The map systems had several features in common. These included a heading-up display format, the use of two zoom levels, the use of a white square to represent vehicle position, and the use of a black square to represent trip destination. The heading-up format presents all maps relative to the direction of travel. As currently applied, this format did not require constant adjustments in map position. Instead, consistency was strived for with slight deviations from the current path not affecting the map orientation. The vehicle position was placed near the bottom of the screen to increase the portion of the map displayed ahead of the vehicle. The maps were updated every $\frac{1}{4}$ to $\frac{1}{2}$ mi to reflect changes in vehicle position.

In addition to the basic features, the dynamic map and advanced systems included superimposed color codes to represent congestion. Color codes showed congestion levels on freeways, but surface street congestion was not displayed. The three codes and their definitions were as follows:

Color	*Congestion descriptor*	*Speed range (mph)*
Red	Jammed	0–15
Yellow	Heavy	15–35
Blue	Moderate	35–50

The advanced experimental system contained all the features of the dynamic map system plus the following additional features:

- *Highlighted alternate route designation* with the starting or diversion point always ahead of the current vehicle position;
- *Textual display bar* to right of map which defined congestion conditions, amount of delay and alternate route suggestions; and
- *Auditory instructions* that were designed to reinforce the visual display information.



a) Static Map *b) Dynamic Map*

c) Advanced Map Zoomed Out *d) Advanced Map Zoomed In*

e) Route Guidance, Normal *f) Route Guidance, Recommended Diversion*

FIGURE 1 Navigation displays.

The alternate route designation used green color coding superimposed over the suggested route. In addition, green triangles were used to indicate the direction to follow. The display bar to the right of the map had two states. In the nondiversion default state, the system recommended that the driver stay on the current route. The congestion level ahead, the cause of the congestion, the expected delay time, and the destination arrival time were displayed from top to bottom. When the system recommended a route change, an auditory beep was first displayed. Then, the cause of the congestion changed to display a diversion message and the expected delay time changed to display the distance to the recommended exit. Advanced system auditory messages were used to provide redundant information. At the start of the trip, the system presented the distance to the congestion, for example, "Jammed Congestion—3 Miles Ahead." Then, as the driver approached a recommended exit, the system gave a diversion message, for example, "Alternate Route—¾ Mile Ahead."

The route guidance system was designed as a simple non-map-based system with features similar to Ali-Scout (*5*) and AUTOGUIDE (*12*). The guidance information included an arrow, expected delay time, and a display of distance to diversion. This system also had two states. In the default nondiversion state, the arrow pointed straight ahead. When the system recommended a change, an auditory beep was first displayed. Then, the arrow changed to point diagonally and the distance to diversion was presented. The route guidance system does not give any advanced notice of congestion as the dynamic map and advanced systems do, and basically provides only diversion recommendations.

Driver Decision Making Simulation

The simulation approach taken here has previously been used to measure both driver and pilot decision making (*13–15*). As indicated in the Figure 2 block diagram, this approach uses a PC computer to control slide projectors, an auditory

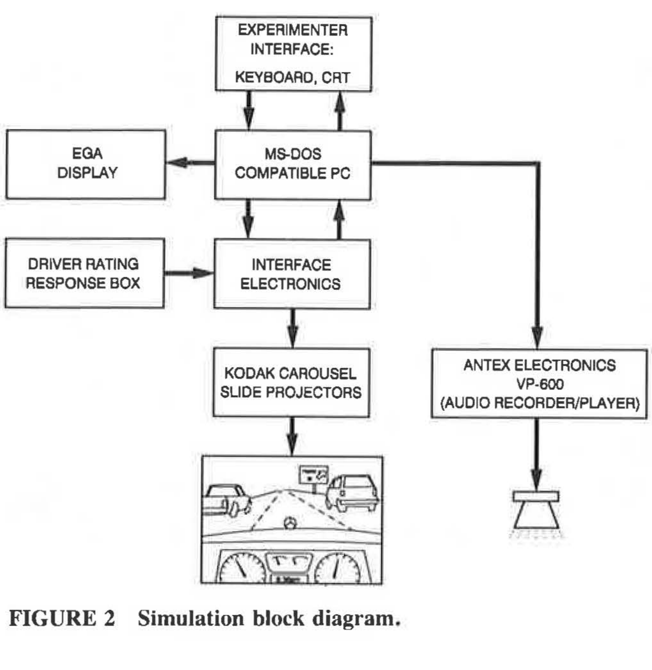


FIGURE 2 Simulation block diagram.

display, and a computer monitor to provide desired stimuli to experimental subjects. In the current application, slides were presented in sequence once every 5 sec showing an out-the-window scene including freeway traffic and guide signs, and a partial instrument panel showing a speedometer, odometer, and digital clock. The slide sequences represented a 10-mi drive west on a southern California freeway (i.e., the Garden Grove Freeway, or Route 22, in Orange County) and represented varying amounts of traffic congestion. Auditory feedback of engine sounds was given that was consistent with the displayed speed. At the same time, a computer monitor presented prototype invehicle navigation displays that were consistent with the congestion scenario.

On the basis of the preceding visual and auditory stimuli, the driver-subject's task was to decide when to divert from the freeway to an alternate route in order to minimize trip delay. To motivate these decisions, driver-subjects were given rewards and penalties according to their performance in minimizing trip delays and in estimating traffic congestion levels in the process of driving to a destination. The reward-penalty structure was designed to simulate real-world motivations, such as saving time, avoiding being late for work, and so on, (e.g., *16*). A summary of the reward-penalty structure as related to trip delay was as follows:

Time Increment (min)	*Reward (saved) ($)*	*Penalty (lost) ($)*
Less than 5	0.00	−0.00
5 to 10	1.00	−1.00
10 to 15	2.00	−2.00
15 to 20	3.00	−3.00
20 to 25	4.00	−4.00
More than 25	5.00	−5.00

The simulation computer kept track of where subjects decided to divert from the freeway route and also queried them about their strategy for returning to the freeway. On the basis of the subject's decision making performance during the driving scenario, the computer then calculated the subject's reward-penalty payoff according to the preceding components. The simulation computer monitor and auditory display were also used for instructions and to present questionnaires to the driver-subjects. The four navigation system visual displays were presented on an EGA color graphics monitor. This provided adequate resolution for displaying street names, street layouts, route guidance symbols, and other textual information. The traffic environment slides were displayed using a Kodak Ektagraphic slide projector. An add-on Tecmar baseboard controlled sequencing and duration of the presented slides. A second slide projector presented slides used for designation of alternate routes.

Auditory stimuli were produced and presented using two add-on boards including an Antex VP-600 which reproduced digitize verbal instructions, and an Adlib sound card which simulated engine sounds representing various vehicle speeds. Specialized keypads were used for answering questions, indicating diversion decisions and designating alternate routes. One keypad included numerical keys, an enter key, and four arrow keys representing up, down, left, and right cursor movements. This keypad was used in lieu of the computer keyboard for answering questions and designating alternate routes. A second keypad with five buttons was used to indicate perceived congestion severity and designate diversion decisions. The physical layout of the simulation is shown in Figure 3.

Driving Scenarios

Several plausible driving scenarios including traffic congestion were needed to provide realistic motivation for use of the prototype navigation systems. Experimental driving scenarios included aspects of traffic incident severity, time constraint, and trip destination. Traffic incident severity involves two factors: (*a*) number of lanes blocked and (*b*) arrival time of the driver relative to the start of the incident. From a traffic engineering point of view, arrival time affects the maximum and minimum average speeds over the travel route in question. It will also affect the distance, or back-up of congestion, from the point of the incident (*17*). To set up realistic congestion scenarios, computer simulation runs were made on the FHWA CORFLO traffic simulation model (*18*) in a scenario that covered the Orange County network, which includes the Garden Grove Freeway. The resulting runs showed that a closure of two-thirds of the available freeway lanes with arrival times of 6:40, 13:20, and 20:00 min after incident occurrence produced the desired delay times.

On the basis of the above traffic flow simulation runs, three incident conditions were defined for the driver behavior simulation consisting of approximately 11-min, 18-min, and 30-min delays based on the freeway speed profiles shown in Figure 4. Road environment slides were prepared that showed traffic congestion consistent with the displayed scenario speeds. Examples are shown in Figure 5. Time constraint or time pressure felt by the driver-subject is affected by trip purpose, amount of time allocated for a specific trip, and time of day. With increasing time constraints imposed on the subject, the level of subject motivation to save time is assumed to increase. The current experiment used commuting to work as the trip purpose. In addition, minimal time was allocated to make the

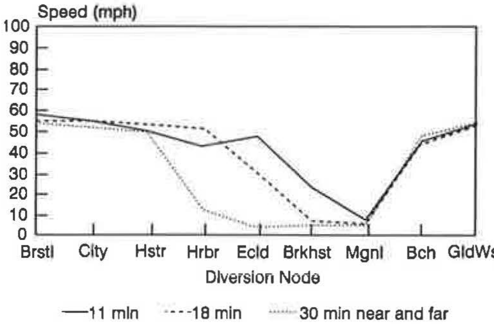


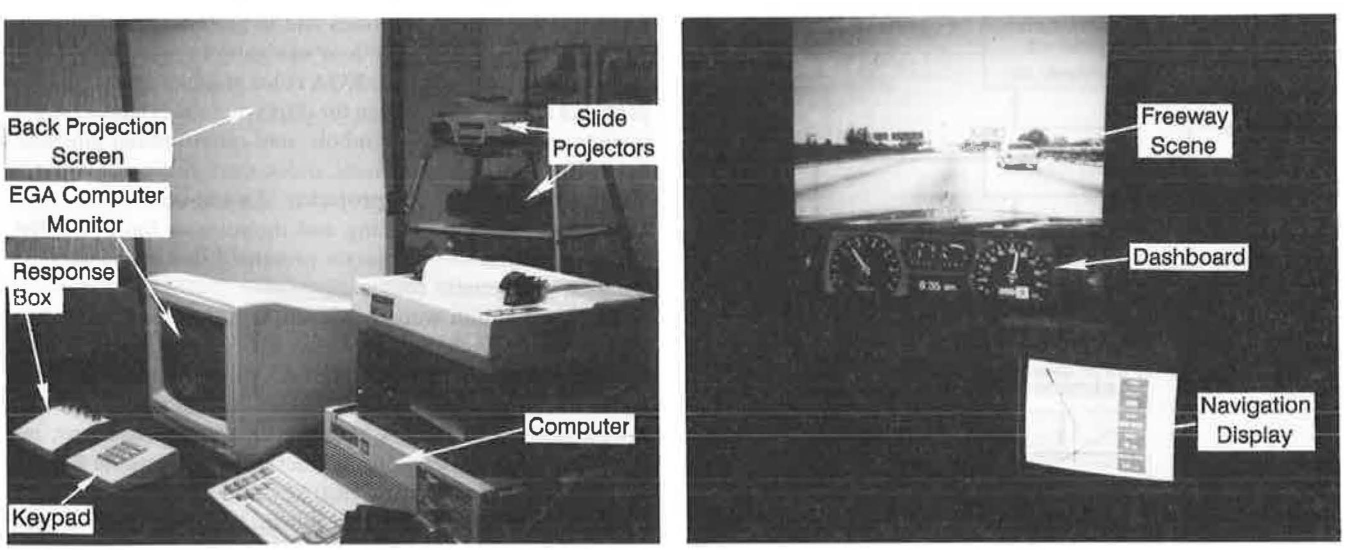
FIGURE 4 Traffic congestion speed profiles.

trip. Thus, a relatively high time constraint was achieved. This time constraint level was applied to all of the scenarios tested.

Trip destination may also have some effect on the percentage of drivers diverting and the selection of alternate routes. For far trips, drivers may see little benefit to leaving the freeway unless encountering severe congestion. For near trips, however, drivers may be more willing to divert from the freeway and go directly to their nearby destination. Most of the drivers in the Orange County network leave the area via one of the freeways. Simulation trips beginning at the Garden Grove Freeway (22 mi) and heading west towards a destination 23 mi away were defined as far trips, while a short trip of 9-mi was defined as a near destination as illustrated in Figure 6. Out of four available scenarios encountered by simulation subjects, three trips involving the 11- , 18-, and 30-min delays were associated with far destinations (23 mi), while a fourth 30-min delay was assigned a near destination (9 mi).

Experimental Design

The experimental design was subdivided into between group and within group factors as follows:



a) Physical Layout *b) In Operation*

FIGURE 3 Simulation physical layout and in operation.

a) Free Flow

b) Moderate

c) Heavy

d) Jammed

FIGURE 5 Road environment slide examples (moderate, heavy, and jammed traffic conditions).

Navigation System (between group)	*Driving Scenario-Congestion Condition (within group)*
None (control)	11-min delay, far
Static Map	18-min delay, far
Dynamic Map	30-min delay, far
Advanced Map	30-min delay, near
Route Guidance	

Each subject-navigation system group was further subdivided as follows:

Age-Background	*Route-Familiarity*	*Gender*
Young (18–29 years)	Familiar	Female
Middle (30–54 years)	Unfamiliar	Male
Old (>55 years)		
Commercial (all ages)		

The basic variable of interest was navigation system configuration. It was felt that a given subject could only be expected to master one system configuration in the limited training time available, so different groups of subjects were assigned to each navigation system condition. Each subject was given all four of the driving scenarios, however. Each of the five subject groups (for each of the navigation system conditions) was broken down into three age groups for noncommercial drivers plus a commercial driving group that included all ages. Driver-subjects were then further categorized according to familiarity with the freeway route and gender.

The sample size of the design totaled 277 drivers taken from various populations. Of the total, 215 were noncommercial drivers and 62 were commercial drivers. Of the 215, approximately 101 were familiar and 114 unfamiliar with the route tested. Finally, an attempt was made to test equal numbers of males and females. Within each gender group, three age brackets were categorized as young (18–29), middle (30–55), and old (> 55). Of the commercial drivers, approximately half were familiar and half unfamiliar. This group was not controlled for gender or age differences.

The majority of subjects were Southern California Automobile Association employees, with unfamiliar subjects recruited at the downtown Los Angeles headquarters and fa-

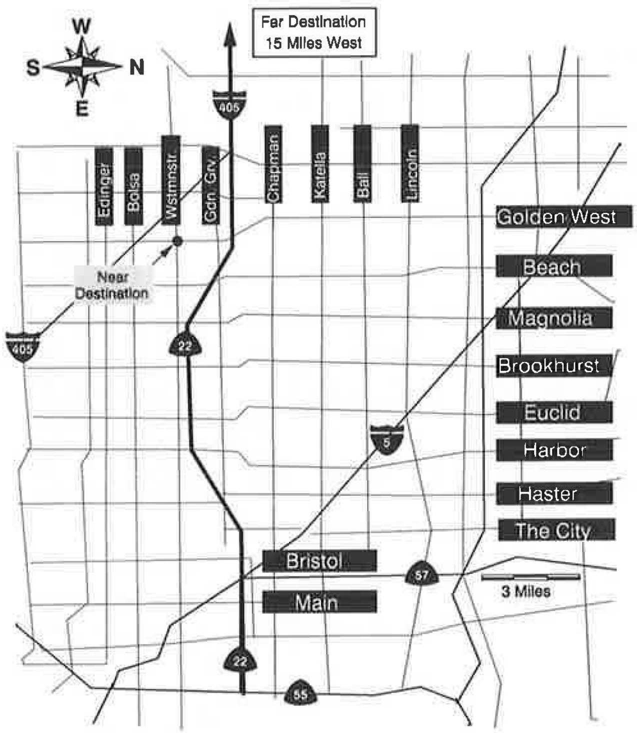


FIGURE 6 Freeway network showing near and far destinations.

miliar subjects recruited from the Costa Mesa processing center near the Garden Grove Freeway route. Additional subjects were recruited from local advertisements and retirement centers to fill out the young and old age categories. Commercial drivers were solicited from airport shuttle services at two locations, one in Santa Ana near the Garden Grove Freeway route and another 30 mi away near the Los Angeles International Airport. The latter location was intended to produce unfamiliar commercial drivers. As a practical matter, it was difficult to obtain commercial drivers unfamiliar with the freeway route and it was also difficult to get familiar drivers in the old-age group. Regardless of the location where subjects were recruited, they were categorized on the basis of answers to a set of pre-experimental questions given verbally by the experimenter.

Experimental Procedures

The experiment was conducted at offices of the Southern California Automobile Club and senior citizens centers where noncommercial driver subjects were obtained and at offices of an airport shuttle service which provided access to commercial drivers. Familiar drivers were obtained at experimental locations near the Orange County freeway network while unfamiliar drivers were obtained at locations 30 mi northwest in the Los Angeles area. Subjects were solicited randomly to the extent possible at each of the locations to fill out the driver demographics required for the experimental design. Upon selection, subjects were randomly assigned to one of the navigation system conditions.

Procedures included orientation, a pre-experiment questionnaire on subject demographics, training, administration of the four driving scenarios discussed previously, and a post-experiment questionnaire regarding typical diversion behavior and attitudes and opinions regarding navigation systems. The orientation, questionnaires, and training for the experimental tasks were automatically administered by the simulation computer to ensure uniformity of presentation. A voice reproduction system was used to administer verbal instructions regarding use of the keypad entry device and interpretation of the visual scenes. Experimental training included familiarization with the visual scenes, navigation system conditions, and the reward-penalty structure. Two practice trials were given, the first without a navigation system for all groups. For the noncontrol groups, orientation was given to the appropriate navigation system. A second practice trial then involved use of the navigation system. The training trials involved exposure to a nonsense scenario involving letter designated offramps (e.g., A St., B St., etc.) to avoid giving any knowledge of the Orange County network to the unfamiliar group.

Formal experimental trials on the Orange County network were preceded by a slide that defined the driving scenario (time of day, trip purpose, and destination). Subjects were then started on the actual scenario which commenced with slides of on-ramp entry, then proceeded with actual Orange County network scenes including appropriate off-ramp signing. Appropriate traffic congestion was portrayed for each of the congestion conditions. Order of presentation of the four congestion conditions was varied across subjects to avoid biasing any condition because of experimental experience. Data collection was automatically performed by the computer in all phases including questionnaire responses and test results. Questionnaire responses, traffic level estimations, diversion decisions, and subject feedback data were all stored in separate data files. Separate files were also used for each subject. At the conclusion of the testing the data files were combined in a spreadsheet for overall analysis. Borland's Quattro Pro (*19*) and Harvard Graphics (*20*) were used for summary analysis and plotting. Stats+ (*21*), which can read data directly from a spreadsheet format, was used for statistical analysis.

RESULTS

The main experimental effects are summarized in Figure 7. The effect of congestion condition on subject diversion patterns summed across navigation systems is illustrated in Figure 7a. Chi-squared analysis shows congestion condition × diversion node to have reliable differences ($p < .001$). The cumulative percentage distributions for diversion show the 30-min delay conditions to be much different than the 11- and 18-min delay conditions. Drivers diverted much earlier for the 30-min delay conditions. The 11-min delay condition also shows a lower overall total diversion percentage (about 84 percent) compared with the other three delay conditions (95 to 97 percent).

The overall effect of age and commercial group subject categorization on diversion patterns is summarized in Figure 7b. Chi-squared analysis indicates the differences to be reliable ($p < .001$). The cumulative diversion distribution shows that the young and middle-aged noncommercial drivers are about the same as commercial drivers in their diversion pat-

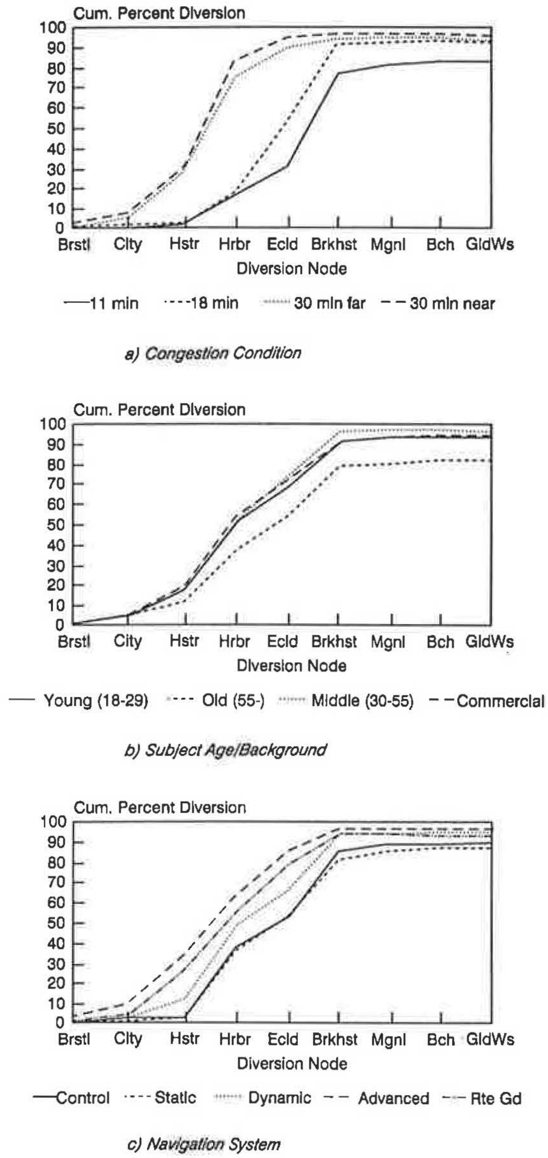


FIGURE 7 Effect of main experimental variables on diversion.

tern, and that the old noncommercial drivers provide the basic difference. The old drivers are more reluctant to divert, with more than three times as many drivers refusing ultimately to divert. Neither route familiarity nor gender had a statistically reliable influence on the diversion results.

The general effect of navigation system on driver diversion was statistically reliable ($p < .001$) and significant as illustrated in Figure 7c. The cumulative distribution shows that the more sophisticated navigation systems allow more anticipation of the congestion condition and ultimately give higher diversion percentages. The advanced system is clearly the best, allowing for the greatest anticipation of the congestion condition and the highest overall diversion rate. The static map condition is basically no different from the control (no navigation system) condition, which is not surprising since it gives no congestion of routing information. The route guidance system gives results similar to the advanced system, the static map is comparable to the control condition, and the dynamic map condition falls somewhere in between.

The advanced and route guidance systems can be compared for compliance with the presented freeway diversion recommendations as summarized in Figure 8. These navigation systems recommended diverting at Euclid for the 11- and 18-min delay conditions and Haster for both 30-min delay conditions. If the subject did not divert at the first recommendation the system recommended diverting at the next exit. To interpret the results, the prevailing speed of the vehicle, the actual level of traffic, the type of navigation system information presented, and subject expectations must be taken into account. For the 11-min delay condition (Figure 8a) the advanced system showed 33 percent diversion at the first recommended exit while the route guidance showed 23 percent diversion. The apparent difference can be explained by noting that the prevailing speeds (Figure 4) were 40 to 50 mph up the point of recommended diversion and only the advanced system subjects received explicit warning of upcoming heavy congestion. The result is a larger number of advanced system subjects diverting before and at the first recommended exit. Generally, the more sophisticated systems produce a greater degree of compliance.

The 18-min delay condition showed about the same 10 percent difference between the advanced and route guidance systems for diversion at the first recommended street (Figure 8b). However, because the prevailing speeds were lower (15 mph), the overall compliance at the recommended node is higher (about 50 percent). The difference between systems may be explained by noting that the advanced system resulted in about 10 percent of subjects diverting at Haster, long before they actually encountered congestion. The cumulative distribution shows equal percentages of subjects diverting by the time they pass Euclid, the recommended exit.

The 30-min delay condition with near destination (Figure 8c) shows 50 percent and 38 percent diversion ratios at Haster, for the advanced and route guidance systems, respectively. Here, the prevailing speeds are relatively high (50 mph in Figure 4), the delay is relatively high and the destination is relatively close. Therefore, the expectation is that drivers will be compliant and divert. The results indicate that the advanced system, which displays the congestion levels and possible alternate routes, may be encouraging more drivers to divert at the first recommended exit than the route guidance system. The redundant verbal messages, the visual coding of traffic, and the textual information of the advanced system give the subject many more cues to upcoming traffic conditions than the simple delay time shown on the route guidance system.

The 30-min delay condition with far destination (Figure 8d) shows only a 5 percent difference between the route guidance and advanced systems at the first recommended exit. The cumulative distributions for these two systems are virtually identical, indicating that the far destination may give the advanced system subjects no greater motivation to divert earlier than the route guidance system subjects. The differences between the advanced and route guidance systems with regard to strict compliance at the first recommended exit may be modified by considering more liberal definitions of compliance. For example, subjects could be considered compliant

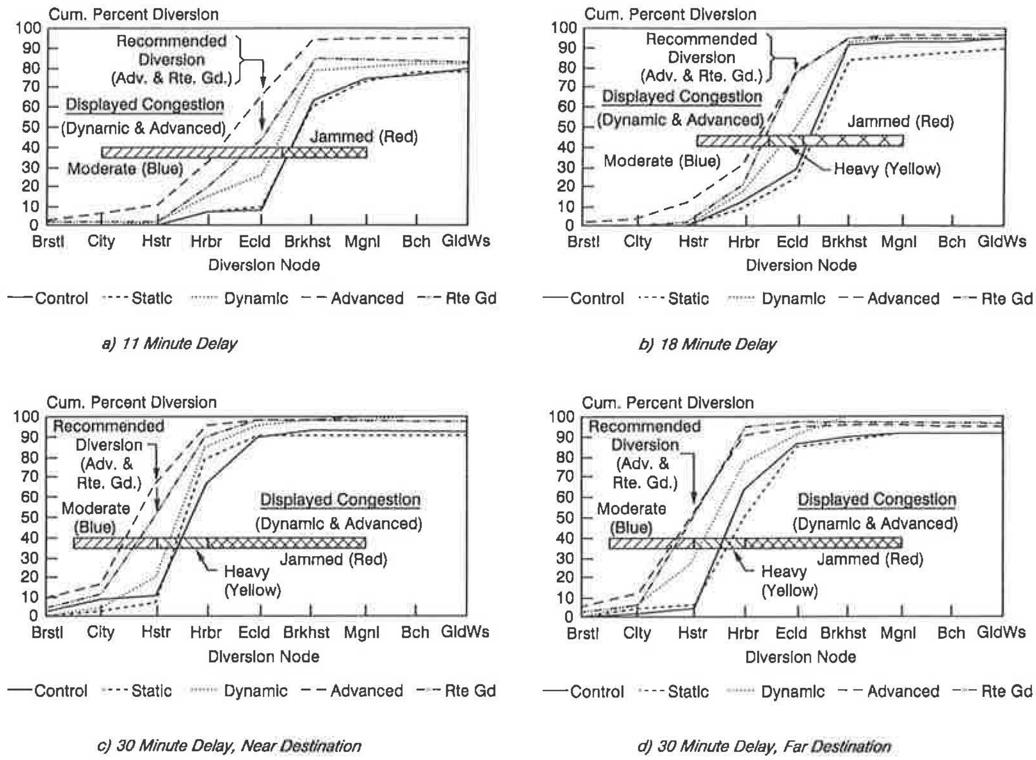


FIGURE 8 Effect of navigation system on diversion distributions.

if they diverted at either the first or second recommended exit.

DISCUSSION

The subject population for this experiment was relatively diverse and probably reasonably representative of typical commuters. Subjects related reasonably well to the simulation and experimental procedures. Diversion response to the congestion conditions as compared to the speed profiles seems quite rational. Of all the subject grouping variables, only age seemed to have any consistent effect on diversion behavior, with the old age group (>55) being more hesitant to divert than younger subjects. Interestingly enough, route familiarity did not seem to have a bearing on route diversion behavior. In this experiment unfamiliar drivers were not any more reluctant to divert than familiar drivers. This could suggest that drivers are comfortable in general with southern California driving conditions so that knowledge of a specific area is not critical. It is also possible that many familiar drivers, although familiar with the Garden Grove Freeway, were not familiar with its environs so that the familiar and unfamiliar populations may not have been significantly different.

Navigation system configuration influenced diversion decisions for all congestion conditions. The static map proved to be no better than the control condition (no navigation system), which is not surprising because the static map gave no feedback on traffic congestion. The advanced and route guidance conditions gave the best results, which is consistent with their navigational capability. The dynamic map system does give feedback on congestion conditions, but offers no route guidance assistance, and so gave performance that was worse than the advanced and route guidance systems but better than the static map and control configurations. This result suggests that a static map system is of marginal help in deciding when to divert from the freeway, although once diverted it would assist in route finding. The route guidance system proved to be nearly as good as the advanced system. For people that have some facility for using maps, a map-based system might be better, again because subsequent to diversion, the map-based system could provide further help in route finding. Computer and display technology is developing at such a rapid rate that a map-based system may not be any more expensive than a route guidance system, and both display formats could be easily provided using the same basic set of information.

The diversion rates for all of the navigation system conditions (including the control or no system condition) were quite high indicating significant aversion to congestion, high compliance with navigation system recommendations, or both. Since the emphasis in this experiment was on diverting to avoid congestion, it is possible that subjects were overly motivated in their diversion response. However, if navigation systems become popular and traffic control management systems are considered to be reliable, it is probable that commuters will have a similarly high motivation for route diversion. For the purposes of subsequent traffic flow analysis of the consequences of driver route diversion, it is possible to scale the cumulative diversion distribution plots to account for lower or higher diversion motivation (e.g., to account for trip purpose, confidence in traffic management system, and so on). Subjects did respond that other conditions, such as trip purpose and certain environmental conditions, would cause lower diversion tendency (or tolerance for longer delays) so a basis does exist for scaling diversion rates. Scaling the cumulative distribution functions by multiplicative factors to vary the effect of subject diversion motivation is suggested. The old-age group effect on the diversion distributions can be considered approximately as a multiplicative effect, and this subject group result should be maintained in further analysis.

REFERENCES

1. G. F. King. Driver Performance in Highway Navigation Tasks." In *Transportation Research Record, 1093*, TRB, National Research Council, Washington, D.C., 1986, pp. 1–11.
2. *Route Guidance and In-Car Communication Systems*, Organisation for Economic Cooperation and Development, Paris, 1988.
3. R. Dial. The ETAK Digital Map Database: Potential Application in Highway Transportation. In *Technology Options for Highway Transportation Operations*. University of Calif., Berkeley, Institute for Transportation Studies, UCITS-P-87-1, 1987, pp. 82–93.
4. R. von Tomkewitsch. The Traffic Guidance System: Ali-Scout. In *Technology Options for Highway Transportation Operations*, University of California, Berkeley, Institute for Transportation Studies, UCITS-P-87-1, 1987, pp. 130–142.
5. R. W. Allen, D. A. Ziedman, and J. F. Torres, *Laboratory Assessment of the Potential Traffic Benefits of In-Vehicle Navigation and Information Systems: Task A Working Paper—Technology Assessment*. WP-2372-1, Systems Technology, Inc., Hawthorne, Calif. Jan. 1989.
6. G. W. Evans. Environmental Cognition. *Psychological Bulletin 88*, pp. 259–287.
7. J. R. Anderson and G. F. Bower, *Human Associative Memory*. V. H. Winston, New York, 1973.
8. S. M. Kosslyn. Information Representation in Visual Images. *Cognitive Psychology*, Vol. 7, 1975, pp. 341–370.
9. L. A. Streeter and D. Vitello. A Profile of Drivers' Map-Reading Abilities. *Human Factors*, Vol. 28, No. 2, April 1986, pp. 223–239.
10. W. B. Verwey and W. H. Janssen, *Route Following and Driving Peformance with In-Car Route Guidance Systems*. Report No. IZF C-14, Institute for Perception TNO, The Netherlands, 1988.
11. W. W. Weirwille, M. C. Hulse, T. J. Fischer, et al. *Effects of Variations in Driving Task Attentional Demand on In-Car Navigation System Usage*. Department Report No. 87-02, IEOR Virginia Polytechnic Institute and State University, Sept. 1987.
12. D. J. Jeffery, K. Russam, and D. I. Robertson. Electronic Route Guidance by AUTOGUIDE: The Research Background. *Traffic Engineering and Control*, Vol. 28, No. 10, Oct. 1987, pp. 525–529.
13. R. W. Allen, A. C. Stein, D. Hill, et al. *Traffic Signal Display Complexity*, TR-1204-1, Systems Technology, Inc., Hawthorne, Calif. March 1986.
14. A. C. Stein, Z. Parseghian, R. W. Allen, et al. *Overhead Guide Sign Visibility Factors, Vol. I: Final Report*, FHWA-RD-88-196; Vol. II: Appendices, FHWA-RD-88-197, U.S. Department of Transportation, Feb. 1989.
15. E. S. Krendel, R. W. Allen, Z. Parseghian, et al. *Pilot Risk Assessment of Windshear Hazards*. TR-1246-1, Systems Technology, Inc., Hawthorne, Calif., Oct. 1988.
16. A. C. Stein, R. W. Allen, and S. H. Schwartz. Use of Reward-Penalty Structures in Human Experimentation. *14th Annual Conference on Manual Control*, NASA CP-2060, NASA Ames Research Center, Moffett Field, Calif., Nov. 1978, pp. 167–178.
17. J. E. Baerwald, M. J. Hubber, and L. E. Keefer. *Transportation and Traffic Engineering Handbook*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976.
18. M. Yeldin, M. Lieberman, B. Andrews, A. Rathi, J. Torres. *TRAF User Guide*. Contract DTFH61-85-C-00005, KLD Associates, Inc., Huntington Station, N.Y., Mar. 1988.
19. *Quattro Pro*, Borland International, Inc., Scotts Valley, Calif., 1989.
20. *Harvard Graphics*, Version 1.01, Software Publishing Corp., Mountain View, Calif., 1987.
21. *Stats+*, Statsoft, Inc., Tulsa, Okla., 1988.

Publication of this paper sponsored by Task Force on Advanced Vehicle and Highway Technologies.