Development Of A Dynamic Equilibrium Assignment Procedure For Network-level Analysis Of New Technology

Permalink
https://escholarship.org/uc/item/9ns859v6

Authors
Sullivan, Edward C.
Wong, Sun

Publication Date
1989
Institute of Transportation Studies
University of California at Berkeley

PROGRAM ON ADVANCED TECHNOLOGY
FOR THE HIGHWAY

Development of a Dynamic Equilibrium Assignment Procedure
for Network-Level Analysis of New Technology

Edward C. Sullivan
Sun Wong

PATH Research Report UCB-ITS-PRR-89-4

Prepared in cooperation with the State of California, Business
and Transportation Agency, Department of Transportation.

The contents of this report reflect the views of the authors who
are responsible for the facts and accuracy of the data presented
herein. The contents do not necessarily reflect the official views
or policies of the State of California. This report does not con-
stitute a standard, specification, or regulation.

August 1989
This paper has been mechanically scanned. Some errors may have been inadvertently introduced.
PATH Goal Statement

The research described in this report is part of the Program on Advanced Technology for the Highway (PATH). PATH research is being conducted at the Institute of Transportation Studies at the University of California at Berkeley, to develop more effective highways. The aim of PATH is to increase the capacity of the most used highways, to decrease traffic congestion, and to improve safety and air quality. PATH is a cooperative venture of the automobile and electronic industries, universities, and local, state, and federal governments.
TABLE OF CONTENTS

EXECUTIVE SUMMARY ........................................... V

I: BACKGROUND .................................................. 1
   1.1.0 INTRODUCTION ........................................ 1
   1.2.0 PROBLEM DEFINITION .................................. 2
   1.3.0 METHODOLOGY ......................................... 3
   1.4.0 MODEL CALIBRATION AND VALIDATION ................... 11
   1.5.0 SUMMARY OF QUESTIONS WHICH CAN BE ADDRESSED THROUGH THIS APPROACH .................................. 12
   1.6.0 CHAPTER CONCLUSION .................................. 16

II: REVIEW OF AVAILABLE COMPUTER SOFTWARE .................. 18
   11.1.0 INTRODUCTION ........................................ 18
   11.2.0 THE SOFTWARE REVIEW PROCESS ....................... 19
   11.3.0 MAJOR EMPHASIS IN THE EVALUATION ................... 21
   11.4.0 RESULTS OF THE REVIEW ................................ 25
   11.5.0 CHAPTER SUMMARY AND CONCLUSION .................... 33

III: LINK COST FUNCTIONS INCORPORATING EXPLICIT QUEUING .... 36
   111.1.0 INTRODUCTION ....................................... 36
   111.2.0 ASSUMPTIONS ......................................... 38
   111.3.0 DERIVATION OF THE DYNAMIC LINK COST FUNCTION .... 39

IV: MODELING FRAMEWORK ........................................ 55
   IV.1.0 DATABASE .............................................. 55
   IV.2.0 NETWORK GENERATOR .................................... 57
   IV.3.0 EXECUTIVE PROGRAM .................................... 57
   IV.4.0 THE DYNAMIC EQUILIBRIUM ASSIGNMENT ALGORITHM .... 58

APPENDIX A ..................................................... 69

REFERENCES AND BIBLIOGRAPHY .................................. 77
iv

LIST OF FIGURES

FIGURE 1: Link Capacity Function ...................... 40
FIGURE 2: Case I Link Queuing Diagram ............... 45
FIGURE 3: Link Capacity Function ...................... 47
FIGURE 4: Case II Link Queuing Diagram .............. 48
FIGURE 5: Case III Link Queuing Diagram ............ 50
FIGURE 6: Case IV Link Queuing Diagram ............. 50
Recognizing that many of the new technologies under consideration in the areas of automation, navigation, and control require a system-wide perspective to evaluate their full impacts, this investigation sought to identify a suitable network analysis procedure which would, with acceptable accuracy, quantify the traffic redistribution effects of deploying new technologies in actual urban settings. The specific focus of the work was to identify a network traffic assignment procedure to be used in the analysis and evaluation of new technologies deployed in selected corridors containing High Occupancy Vehicle (HOV) facilities; however, the results are equally applicable to the assessment of new technologies in any congested urban network.

Based on the nature of urban traffic and the likely characteristics of the technologies to be deployed, and because of the need for more precision than normally is obtained with the network assignment models typically used for transportation systems planning, the following criteria were developed:

CRITERION 1. The network traffic estimation procedure should be an equilibrium assignment method, capable of estimating either a user-optimum traffic pattern or a system-optimum pattern.
DISCUSSION: In this report, "link cost function" refers to the relationship between the travel time (or cost) on a link and the link traffic flow, expressed in vehicles per hour. Equilibrium assignment methods, which use monotonically increasing link cost functions, are commonly used for traffic estimation in urban networks. A user-optimum assignment, where no trip can improve its individual travel time (cost) by diverting to an alternate route, is an equilibrium assignment obtained by setting the link's cost functions to the average cost functions for the links (that is, the total link time/cost functions divided by the flow). By contrast, a system-optimum assignment, where total travel time (cost) is minimized, is an equilibrium assignment obtained by setting the link cost functions to the marginal link cost functions. Therefore, in principle, readily available equilibrium assignment procedures can be used to estimate either user-optimum or system-optimum traffic patterns; however, some network performance measures, such as travel times and delays, are calculated assuming that results represent user-optimum patterns, and would need to be recalculated to describe system-optimum results.
CRITERION 2. In addition to user-optimum and system-optimum flow patterns, the traffic estimation procedure must be able to estimate constrained system-optimum patterns, in which individual routes are not permitted to exceed some multiple of the corresponding user-optimum travel times (costs).

DISCUSSION: The fundamental problem with system-optimum assignments is that total network time or cost may be minimized by forcing a minority of trips to follow extremely indirect and costly routes, in order to benefit the majority. Clearly, such patterns are not feasible in a system where the cooperation of individual drivers is a necessity, nor is such a solution equitable. On the other hand, if total travel time (cost) in the system could be reduced significantly by diverting some vehicles to slightly longer paths than their shortest available paths, the difference would be insignificant for the individual travellers but could result in substantial savings elsewhere in the system. The development of traffic control strategies based on the constrained system-optimum assignment principle seems important in order to realize the full benefits of certain new highway technologies;
therefore, the availability of an analysis procedure to evaluate such strategies is fundamental.

CRITERION 3. The traffic estimation procedure must be able to deal explicitly with queuing and represent the delays due to queues, in a dynamic analysis framework.

DISCUSSION: A large amount of the travel time and cost accrued in congested urban networks results from queues behind bottlenecks. The network analysis procedures commonly used in transportation systems planning approximate queue delays through the extension of the link cost functions to flow values beyond capacity. But, this approximation ignores the dynamic nature of queuing, and the fact that queue lengths and delays are more sensitive to the time distribution of demand than to the total amount of demand accommodated during an entire peak period. The only way to make a reasonable approximation of queuing is to perform a dynamic analysis, by dividing the entire analysis period into fairly small time increments, as is done with models used for traffic operational analysis, such as FREQ. Since the application for this procedure is network planning, the level of modelling detail and the
amount of calculating effort involved should be less than in the models used for sub-network operational analyses. However, it is clear that estimates of the starting time distributions for trips is essential in order to deal adequately with arrival demand distributions at bottlenecks.

CRITERION 4. The procedure should deal explicitly with the fact that, in a multiple time period analysis, some trip lengths are long relative to the length of the time increments considered: thus, individual trips can appear on the network in more than one time increment.

DISCUSSION: To deal with this situation, it is clear that the procedure must be able to estimate how much of a long trip is completed during each time increment and then adjust the origin-destination matrix for the next time increment to show the incompleted portion of that trip originating at an appropriate enroute location. Since trips obviously do not all start at the beginning of time increments, it is also necessary to deal with the distribution of start times, which generally should be considered uniform within each time increment (except, of course, for the trips which began during previous
CRITERION 5. Recognizing that the required traffic assignment procedure is intended for application to large portions of urban areas, data requirements for representing individual transportation facilities must be limited, and must be compatible with the data collection capabilities of typical transportation planning organizations.

The study approach first involved a state-of-the-art review of selected urban transportation planning and traffic simulation computer programs, in order to determine whether one or more of the available software packages met our criteria, or came close enough that the dynamic equilibrium assignment procedure could be
developed as an extension of one of the existing packages. The state-of-the-art review considered the following software packages: MINUTP, TMODEL2, TRANPLAN, CARS, MicroTRIPS, EMME2, FREQ, and NETSIM. Only software packages for IBM-compatible microcomputers were included. Not surprisingly, none of the existing packages exactly met the criteria; however, it was decided that the new dynamic equilibrium assignment routine should be implemented as an extension to an existing planning package, in order to take advantage of available network and matrix manipulation capabilities, and additional modeling features. Because of familiarity and its availability at the ITS, we feel comfortable recommending that the dynamic equilibrium assignment model be made compatible with the MINUTP data protocols: however, in principle, TRANPLAN, EMME2, or one of the other available packages might equally well be used.

Recognizing that to satisfy the above criteria requires the development of a new equilibrium assignment procedure, attention was given to the characteristics of such a procedure. An equilibrium assignment algorithm was designed which can be used in a dynamic (multiple time increment) framework, in which the link cost functions are adjusted before each time increment to represent the link travel times and queue delays which would actually be present during that period. The algorithm explicitly considers the time distribution of demand across and within the increments, and carries over the portions of trips which are not complete at
the end of each time increment so that they become part of the demand for the following time increment. The data structure and network generation requirements to implement such an algorithm are considered.

A most interesting aspect of the proposed procedure is the explicit treatment of queueing required to generate appropriate link cost functions for the equilibrium assignment applied in each time increment. The approach begins with a "base link cost function," which is used whenever no queue is present, and which is equivalent to the top portion of the empirical speed-volume relationship for the facility in question. Then, four cases are identified: (1) when a queue is formed sometime during the time increment, (2) when a queue exists throughout the time increment and is growing, (3) when a queue exists throughout the time increment and is shrinking, and (4) when an existing queue disappears sometime during the time increment. Queues which are formed and disappear entirely within one time increment are considered transient and are ignored, in keeping with the assumption that demand rates are constant in each time increment. (Time increments are fairly short, on the order of five to thirty minutes.) It turns out that the link cost functions derived for these four cases, as well as the base cost function, are all special cases of a single generalized cost curve, which is a function of link capacity, demand, time increment length, and queue length at the beginning of the time increment. Because the equilibrium assignment procedure requires that link cost functions
equilibrium assignment procedure requires that link cost functions be monotonically increasing for all positive values of link flow, it was necessary to develop an approximation for Case 4, since the actual cost function in that instance is U-shaped.

This report carries the development of the dynamic equilibrium assignment procedure to the end of the conceptual and derivation stages. The next step is to implement the procedure, test its ability to estimate traffic patterns in actual networks (under the assumption of user-optimum driver behavior), and then apply the procedure to address some of the researchable questions raised in this report.

The central recommendation of this report is that, following review of these concepts and their relationship to PATH needs, the new dynamic equilibrium assignment procedure be implemented as the basis for network-level technology evaluations to be performed within this research program, including the systems analysis task of the HOV Feasibility Study, through which the development effort to date has been funded.

The body of the report contains four chapters. Chapter I is a detailed discussion of the motivation for this work, the thinking behind the selection criteria, and representative questions related to new technology deployment which the proposed dynamic equilibrium assignment procedure would help to answer. Chapter II contains the
review of existing software packages, and summarizes their features which are pertinent to the current investigation. Chapter III addresses in detail the derivation of link cost functions in a way which deals explicitly with queuing. Chapter IV addresses the overall modelling framework within which these link cost functions would be used, discusses data requirements and data processing procedures, and presents the detailed specification of the dynamic equilibrium assignment algorithm.
I: BACKGROUND

1.1.0 INTRODUCTION

In coordination with other research projects of the ITS Program on Advanced Technology for the Highway (PATH), this research has focused on identifying the potential benefits of improving traffic network performance, using advanced communication and control systems. The objective is to provide suitable network analysis techniques, which are essential to evaluate the feasibility of introducing advanced highway technologies in urban settings, such as on high occupancy vehicle (HOV) facilities operated independently or in conjunction with "Smart Corridor" installations. The analysis techniques will permit effective investment and control strategies to be sought, in light of both unpredictable incidents and recurrent congestion conditions, with proper consideration of network-wide implications.

The specific intent of this work is to select or design a network traffic assignment model which will help determine network control parameters such as adjustments in link capacities and driver guidance procedures, based on estimation of network performance characteristics including queuing, delays, and flows. Based upon these parameters, the model should be useful in generating alternative technology deployment and operational control strategies to improve network efficiency for a wide range of hypothesized operating conditions.
The ultimate objective of the research is to evaluate and be able to apply the model to any real-world network.

1.2.0 PROBLEM DEFINITION

If a link or links within an urban road network experience temporary closure or reduced capacity due to some incident, such as a vehicle breakdown or accident, or if they simply experience recurrent congestion during certain time periods, this information should be able to be passed to up-stream vehicles approaching these links, as well as to other vehicles in the network. Under such conditions, being able to divert certain traffic to alternative travel routes before becoming trapped in the queues leading to the problem links may reduce further worsening of congestion on these links and therefore may reduce the total delay experienced by the vehicles in the system. It is important to realize that the trip diversion strategy may be applied not only to the traffic intended for the links experiencing problems, but also to other traffic, in order to improve the overall network operating efficiency. Existence within the network of certain special links, like HOV facilities equipped with some form of advanced traffic communication and control technologies, perhaps with excess capacities and/or installed capacity augmenting technologies, adds to the complexity, as well as the opportunity for improvement of the general traffic network congestion problem.
Patterns of traffic diversion should be estimated by the model based on the individual link control possibilities (determined by technology) and flow conditions. Such diversion can be influenced by the advanced communication and navigation technologies which will be assumed to be available at some time in the future. The consequences of different technologies, and different levels of geographic and market penetration of these technologies within the overall urban transportation system can be explored, in order to help evaluate both the benefits of the technologies and their alternative deployment strategies.

1.3.0 METHODOLOGY

Vehicle assignments to alternative travel routes for any traffic control strategy can be tested using several different methods. The methods vary along two significant dimensions: whether flows are considered in a single or multiple time period framework, and which principle is employed as the objective function. With reference to the latter, it is intended that the method provide solutions located along a continuum ranging from user-optimum to system-optimum assignments. A final significant issue has to do with the representation of travel delays, especially the delays due to queuing which have never before been adequately addressed in assignment procedures used for the analysis of large scale urban networks.
The issue of whether single or multiple time periods are considered is discussed first. Further attention is then given to the nature of the objective function and constraints. Finally, the need for better dynamic representation of delays due to queuing is discussed with reference to the appropriate level of detail to implement in a planning model.

1.3.1 Single-Time Period Assignment

Under link incidents or oversaturated conditions, alternative travel routes for the traffic on a link will be identified and certain percentages of these vehicles assigned to alternative routes connecting their original origins and destinations. The assignment will be based on the average flow conditions on the alternative routes at the assignment time, based on whichever travel time optimization criterion is being applied. By diverting to the alternative routes, the anticipated delays should be less than the expected delays on the links experiencing the incident or recurrent congestion conditions.

However, this method, which is typical of UTPS-like trip assignment procedures, does not provide an opportunity to examine the dynamics of traffic operations within the network over an extended period. By applying the above single-time period assignment method, there is no consideration given to the changes in flow conditions on the links due to changes in demand across time. Furthermore, trips making route choices based on current conditions in the trip
assignment may later face altogether different conditions in the down-stream traffic. The single time period assignment method is not able to consider these dynamic aspects of the problem. Acknowledging the dynamic aspects is necessary, especially when time-variant control strategies are sought.

1.3.2 Multiple Time Period Assignment

Multiple time period assignments to alternative travel routes can be based on the same routing criterion as in single time period assignments. However, the link flow conditions and route choices with respect to travel time are analyzed for individual time increments, like 10 to 30 minutes. Based on the link flow conditions determined in each time increment, the trip assignment is conducted to achieve an equilibrium travel pattern throughout the network, based on the selected optimization criterion.

Recognizing that long trips may span two or more time increments, especially if congestion is heavy, trips which will not reach their destinations within a single time period travel only as far as an intermediate pseudo-destination, located along the travel route determined by the selected routing criterion. The demand matrices for trips starting in each time period are increased by the remaining portions of the trips which began in the previous period, but progressed only as far as an intermediate node by the beginning of the time period in question. This trip assignment approach
provides the opportunity to detect changing conditions after each

time increment and re-establish the equilibrium within the network.

It is clear that the multiple time period assignment method has the

advantage of being able to consistently track the network-wide

conditions through an extended period much more precisely than the

single period assignment method. On the other hand, the approach

is still fundamentally a planning model, involving more averaging

and approximation than the microscopic simulation models

characteristic of local area traffic operations analysis.

In order to implement either the single or multiple time period

network planning models, a suitable computer program with an

efficient equilibrium assignment algorithm is necessary.

1.3.3 Objective Function and Constraints

In most transportation planning and traffic operations analyses,

the fundamental optimization criterion is minimizing travel time.

This research project is no exception, although the measure of

performance will be based on different objectives.

One objective is "User-Optimum", which is based on the principle

that average user costs (travel times) on alternative travel routes

being used between each origin-destination pair are equal. This

method is simply the pursuit of minimizing individual user's

perceived travel costs. Based on this approach, users who divert
to alternative travel routes directly benefit in their perceived travel time, but the original users of these routes may be adversely affected by the increased traffic by experiencing longer delays, which typically leads to an increase in total system travel cost. Therefore, the overall system operating cost generally is not minimized.

An alternative objective is "System-Optimum". This is based on the analysis of alternative routes' marginal costs, not the average costs. The fundamental goal is to distribute traffic so the marginal costs on all alternative routes being used are equal. By applying this criterion, the total system operating cost can be minimized under the prevailing demand conditions. However, some users' travel times may be considerably longer than their travel times under user-optimum conditions.

Both of these optimization approaches can be utilized in conjunction with the single time period and multiple time period trip assignment methods. It is also intended to pursue intermediate objectives, for example, a system-optimum assignment in which the paths used for each origin-destination pair are constrained to not exceed some multiple of the travel time achieved under user-optimum conditions. Control strategies based on such intermediate strategies may provide improved but still practical results in dealing with real-world traffic networks.
Computationally, achieving either a user-optimum or a system-optimum assignment can be accomplished by the same equilibrium-seeking algorithm. To achieve user-optimum, link cost (time) functions must represent user average (perceived) costs; for system-optimum, the link cost functions are simply set to the equivalent marginal costs. In both cases, the average or marginal costs (times) are expressed as functions of link flow, measured in vehicles per hour. Equilibrium assignment algorithms which can be used to estimate both user-optimum and system-optimum flow patterns in a network are readily available through UTPS and several UTPS-like software packages.

Although conceptually straight-forward, to our knowledge there exists today no software capable of estimating a constrained equilibrium assignment, of the type needed to explore the types of intermediate objectives described above. Such an algorithm would need to keep track of both marginal and average link cost functions, and, during system-optimum assignments, systematically exclude from consideration any paths which exceed the stated average path cost constraint. This requires a "next best path" logic, for which algorithms are readily available in the literature. Implementing this logic in a practical software package is included among the recommendations of this report.
1.3.4 Explicit Representation of Queue Delays

Much of the travel time spent in congested urban networks results from queues behind bottlenecks. The network analysis procedures commonly used in transportation systems planning approximate queue delays by extending the link cost functions to flow values beyond capacity. Unfortunately, this approximation ignores the dynamic nature of queuing, and the fact that queue lengths and delays are more sensitive to time distributions of demand than to the total amount of demand accommodated during a peak period, which is usually the only information provided to network assignment procedures used in planning applications.

The only way to make a reasonable approximation of queuing is to perform a dynamic analysis, by dividing the entire analysis period into fairly small time increments, such as done in the FREQ model. Since the application for this procedure is network planning, the level of detail and the amount of calculations involved should be less than in the models used for small network operational analyses. However, it is clear that estimates of the starting time distributions for trips is essential in order to deal adequately with demand arrival time distributions at bottlenecks.

In order to deal adequately with the effects of queues on link cost functions, it is necessary to consider four cases: (1) when a queue is formed sometime during the time increment, (2) when a queue exists throughout the time increment and is growing, (3) when a
queue exists throughout the time increment and is shrinking, and (4) when an existing queue disappears sometime during the time increment. Queues which are formed and disappear entirely within one time increment may be considered transient and ignored, in keeping with the assumption that time increments are short enough that arrival demand rates at each bottleneck are constant within each time increment.

The effect of each case of queueing on the shape of the link cost function is different. It seems clear that the link cost function for Case 1 is equivalent to the top part of the empirical speed-flow relationship for the facility in question, up to capacity (i.e., as long as the flow is less than capacity, no queue delay is involved). Beyond capacity, the link travel time is based on the link speed at capacity plus any queue delay, which depends on the arrival rate, the capacity, and the length of the time increment. Cases 2 and 3 are different, since their link travel times, at all demand levels, are determined by the link speed at capacity plus the queue delay, which depends on the previous parameters plus the queue length at the beginning of the time increment. The last option, Case 4, is the messiest, since average link time involves some portion of the traffic operating at the speed of capacity, and some operating at higher free flow speeds.

In light of the above, it seems clear that it is impossible to precisely define link cost functions for use with an equilibrium
assignment procedure in the absence of at least some knowledge about the time distributions of queuing. This is precisely what the proposed model will do. By permitting multiple time period analysis, it will be possible to keep track of the creation and eventual disappearance of queues in different time periods, and thereby determine the appropriate link cost functions to use in each time period. The precision of this approach will, of course, be directly related to the length of the time periods. The tradeoff will be between the greater precision given by small time periods and both the computational effort and the availability of information about starting time distributions for the origin-destination matrix during the analysis period. The time periods cannot be so short that they stretch the credibility of the available data on trip starting time distributions when applied to individual cells of the O-D matrix.

1.4.0 MODEL CALIBRATION AND VALIDATION

As part of developing the model capability desired for this research, it is important to test the model with real-world traffic data to ensure its applicability in real-world network control applications. In particular, reasonable assignment results should be able to be obtained for an existing urban road system before the model is applied in research directed to evaluating the consequences of new highway configurations and/or the deployment of new technologies.
The test network should be composed primarily of links representing freeways and major arterials. Traffic on these roadways would usually account for the majority of commuting traffic in any metropolitan area. On the other hand, some distortion is introduced as in all network planning models, due to the crude representation of minor arterials and local streets. Generally, the traffic service features of minor arterials are combined into those of nearby parallel major arterials, and the attributes of local streets are modelled by abstract links called "centroid connectors." Providing a framework for calibrating link cost functions, especially those representing surface arterials, in a way that reduces the characteristic distortions of planning models when applied to traffic management, is another notable objective of the current modelling effort. This issue is addressed in greater detail in Chapter III of the report.

1.5.0 SUMMARY OF QUESTIONS WHICH CAN BE ADDRESSED THROUGH THIS APPROACH
Based on the problem and methodological issues described in the preceding sections, we can identify the following list of specific questions which can be addressed in this research. These questions relate to both analytical issues and to forming strategies for the deployment of advanced technologies in the real world.

1. Assume that a percentage \textit{"X"} of the traffic has access to advanced communication and navigation technology to know at all times the "best" route to take toward their destinations.
(Here, the definition of "best" is determined by the optimization criterion in effect.) How do the measures of network performance (e.g., total delay, equity in the geographic incidence of delays) vary as "X" ranges from 0 to 100%? Do optimal values of "X" exist, for some optimization criteria, at X < 100%? Note that addressing this question requires that the assignment procedure permit traffic allocation to a preloaded network, where the preloaded volumes represent the (100-x)% of the traffic not equipped with the advanced technology.

2. Equilibrium traffic patterns can be estimated by either user-optimum or system-optimum trip assignment. What is the difference in overall system operating cost between these two optimization criteria, in representative California travel corridors?

3. When network flows are estimated based upon the system-optimum trip assignment criterion, a small number of trips may be diverted to alternative routes with much longer travel times. Under such conditions, a small group of users are heavily penalized and may refuse to follow the directions given by the traffic control system. This would, of course, adversely affect the total travel time within the network. However, if the penalty can be constrained to be evenly distributed among a larger group of users, the penalty on each affected user's
travel time can be made relatively small, and users should be willing to follow the alternative, slightly longer routes. What would be the effect of following such a near-system-optimum assignment criterion with regard to the system measures of performance, in comparison with both the true system-optimum and true user-optimum conditions?

4. Impacts of traffic incidents vary depending on their locations and durations as well as on the effectiveness of the incident control procedures. If an incident persists for a relatively long period, then advanced communication and navigation technologies should be increasingly effective in reducing the overall travel time within the network. However, if the incident persists for only a short period and traffic flow quickly returns to normal, it may be that traffic control actions based on transient conditions may result in making matters worse. Is it possible to establish a clear cut relationship between the duration of incidents and the performance of the communication and navigation technologies, for different optimization criteria?

5. When the multiple time period trip assignment method is used, the uncompleted portions of the trips from the previous time increments will be added to the new demand matrix in each time period. Then a new equilibrium trip assignment will be performed and new travel routes will be identified. Under
such a method, some users may be shown as changing travel paths from their original routings several times. Will such a procedure produce trip routings whose travel times are within specified tolerances of the user optimum times under average conditions for the entire trips? Will the resulting routes appear reasonable, that is, not overly circuitous and erratic?

6. The advanced traffic control technologies of interest include the ability to change operations at signalized intersections dynamically to meet the new traffic demands after diversion. Such changes can include adjusting traffic signal timing and adjusting the number of through and left-turn lanes (perhaps to zero) using changeable signs. What will be the effect of such changes on the users who travel through these locations regularly under non-incident conditions?

7. Due to the introduction of advanced technologies, certain groups of links in a network may have special operating characteristics, such as higher than normal speeds and/or capacities. An example of such links would be an automated freeway, operated as an HOV facility. How does the presence of such facilities, and their placement, affect the performance of the overall network, under both incident and recurrent congestion conditions, for different optimization criteria?
One would expect that advanced technologies would provide different levels of benefits depending on the topology of the transportation network. A priori, it seems grid networks should experience more benefits than corridor-based networks, which in turn should be benefited more than single-spine or hub-and-spoke networks. What are the magnitudes of these benefit levels, for different network forms actually found in California metropolitan areas?

The above questions were identified at the outset of this project as a framework for studying the impacts of different advanced technologies at the network planning level. Clearly, some of these questions are interrelated and should be addressed as such. Also, future investigations may not be limited just to these questions. Additional issues may also be addressed, and the insights gained from early efforts to address these questions as stated may indicate that some other questions need not be considered at all, or that they should be addressed in a different way than stated here.

1.6.0 CHAPTER CONCLUSION
The research objectives, issues, methodologies, and traffic assignment criteria have been identified. Although some aspects of the research are so far speculative due to uncertainties about the future technologies to be available, it seems reasonable to assume that the fundamental traffic flow and equilibrium trip
The assignment principles on which this research will be based are valid. The degree of success in applying the proposed model to accomplish the tasks stated in the problem definition will mainly depend on the availability of suitable computer procedures to apply the methodology and assignment criteria to a particular network.

Rigorous testing of any network assignment procedure is almost always problematic, since it requires accurate information on origin-destination demands, route choices, extensive traffic flow data, and good information regarding link capacities and travel time-flow characteristics. While such information is seldom fully available at the individual facility level, if the model is applied at the regional corridor level, it does appear that adequate data can be obtained. In particular, fairly good origin-destination data for work trips are available from the census, and both facility and traffic volume-speed information is available from the California Department of Transportation (Caltrans) and local government records.
II: REVIEW OF AVAILABLE COMPUTER SOFTWARE

11.1.0 INTRODUCTION
The success of the proposed traffic network analysis model will mainly depend on its capabilities for network traffic assignment and link queuing analysis. The general scope of the desired features are discussed in the previous chapter. Currently, many transportation planning computer software packages already have certain of these features, although the actual capabilities of these packages for solving real world problems vary in their flexibility, data manipulation power, and underlying theory.

Before reaching any conclusions regarding the need for new software development, it was decided to perform a broad review of the capabilities of existing software. If one or more existing packages matched or came fairly close to providing the desired traffic assignment capabilities, then it obviously would be advantageous simply to use or adapt that available software. Even if no existing assignment procedure met this project's needs, there could still be merit in making any new software developed by this project compatible with the data structure of an existing software package, to take advantage of available network and matrix manipulation features and/or ancillary capabilities (like mode choice and trip distribution models).
A review of existing software was performed, using resources already available at the Institute of Transportation Studies (ITS), Berkeley. Of the many different packages currently available in the market, several are installed at the ITS. In addition, many publications concerning the capabilities and past experience using the software packages are also readily available.

The initial comparison among packages mainly employed a literature review, to identify the availability of the desired features in each package. Then, based on the findings of the literature review, several packages were selected for further testing with sample problems. Based on these tests, the package or packages with the most appropriate capabilities for this research can be identified.

11.2.0 THE SOFTWARE REVIEW PROCESS

The software review process established a list of packages likely to be applicable, researched available literature, and tested some packages with sample problems. Several sources were consulted in compiling the review list.

One major resource is the "Software and Source Book" published by the Center for Microcomputers in Transportation, a transportation technology transfer agency established by the U.S. Department of Transportation at the University of Florida. This reference contains introductory information on virtually all of the
transportation planning and traffic operations analysis packages currently available in the market. It was decided to limit the review to microcomputer-based software, both for convenience of use and for ultimate ease of transferability of the software selected. The names and developers of the selected packages are listed in Table 1.

**TABLE 1**

**LIST OF PACKAGES SELECTED FOR REVIEW**

<table>
<thead>
<tr>
<th>NAME</th>
<th>DEVELOPER</th>
</tr>
</thead>
<tbody>
<tr>
<td>MINUTP</td>
<td>Comsis Corp.</td>
</tr>
<tr>
<td>TMODEL2</td>
<td>Professional Solutions, Inc.</td>
</tr>
<tr>
<td>TRANPLAN</td>
<td>The Urban Analysis Group</td>
</tr>
<tr>
<td>CARS</td>
<td>Roger Creighton Asso. Inc.</td>
</tr>
<tr>
<td>MicroTRIPS</td>
<td>MVA Systematica, UK</td>
</tr>
<tr>
<td>EMME</td>
<td>University of Montreal</td>
</tr>
<tr>
<td>FREQ</td>
<td>University of California, Berkeley</td>
</tr>
<tr>
<td>NETSIM</td>
<td>KLD Asso. Inc.</td>
</tr>
</tbody>
</table>

All of this software, except FREQ and NETSIM, are comprehensive transportation planning packages designed to handle highway networks, manipulate trip matrices, and perform traffic assignment. These packages were developed to emphasize broad network investment and capacity-allocation issues, rather than the operating efficiency of specific facilities. FREQ and NETSIM, on the other hand, were developed specifically for highway and urban arterials operational analysis.
An evaluation checklist table was established listing the desired features of the assignment model. This evaluation table was refined as the evaluation progressed to clarify some additional important features which were identified and to highlight some other differences among packages.

The results of the preliminary review, presented in Table 2, are primarily culled from user's manuals for the different packages, from articles published by users, or directly from the package developers through telephone interviews. It should be noted that, first, the evaluation is conditioned by the resources available at the ITS and, second, the evaluation is not comprehensive, but rather is limited to the features of interest. Further, the evaluation results may not be fully up-to-date for certain packages, because the information was beyond the reach of the authors at the time the evaluation was made, nor were the final results commented upon by the developers of packages reviewed. Finally, it should be noted that a review of this nature must deal with a moving target, since most of these packages are continually being enhanced by their developers.

11.3.0 MAJOR EMPHASIS IN THE EVALUATION

Through the review process, two major features were given special consideration due to their importance for the proposed model. One is the trip assignment method, and the other is the link queuing
analysis method. The desired capabilities of these two features are described in detail in the following paragraphs.

11.3.1 Equilibrium Traffic Assignment
As discussed previously, the proposed model will re-distribute some of the traffic in the network from over-congested links to alternative routes in order to reduce delays and improve the network operating efficiency. The trip re-distribution process, which we call trip diversion, requires a trip assignment method to re-establish a new equilibrium state within the network after an incident or other change of condition.

Currently, equilibrium traffic assignment, based on Wardrop's Principle #1, is available in many of the computer packages. This trip assignment method, with suitable preprocessing of the network representation and demand matrices, should fulfill the trip assignment tasks required in the proposed model. However, the algorithms used to implement equilibrium assignment in the different computer packages are different, which implies that the outputs from these models may vary somewhat. These differences are unimportant as long as the algorithm is a correct implementation of the equilibrium principle.

Within a traffic network, users choose their routes to minimize individual perceived travel cost. On the other hand, from a transportation planning perspective, it would be desirable to
allocate the users within the network in such a way as to minimize the total travel cost within the system. The former is the user-optimum equilibrium trip assignment and the latter is the system-optimum equilibrium trip assignment.

The user-optimum equilibrium within a traffic network is thought to represent people's travel behavior, assuming logical decision making and perfect knowledge of conditions. Assuming complete knowledge of their available options, users should always take the shortest path between a particular origin-destination (O-D) pair. But all users would not necessarily take the same path, due to capacity restraints on links between the O-D pair. They may be split among a few equally shortest paths within the network. The equilibrium is reached when the users assigned to different paths can no longer improve their individual travel costs by unilaterally switching to other paths. However, in general, the user-optimum equilibrium does not minimize the total travel cost in the system.

The system-optimum trip assignment method allocates traffic among the paths within the network between different O-D's in such a way as to minimize the total travel cost in the system. Under this form of equilibrium, the marginal costs on all alternative paths utilized between each O-D pair are equal. However, the user's individual travel costs are generally not minimized. It should be realized, as Kanafani stated in "Transportation Demand Analysis," that the system-optimum equilibrium does not represent typical
users' behavior. Rather, it is a tool to analyze the situation when some means of traffic control is contemplated to move the pattern of trips from user optimum toward system optimal.

The importance of being able to implement both trip assignment methods in the proposed model is to be able to generate appropriate system control strategies to improve the overall network operating efficiency. In principle, an equilibrium assignment procedure should be able to produce both, depending on whether link cost functions represent average or marginal costs. In practice, however, some summary statistics (such as total travel time and total delay) are meaningless for system-optimum assignments estimated by programs designed to estimate user-optimum patterns.

11.3.2 Explicit Link Queuing Analysis
In order to realistically model user-perceived delays and determine the quality of the traffic flow within a network, not only must conventional link parameters, such as average delays and average travel time, be determined, but also explicit queuing phenomena must be evaluated. The link queuing estimates are important parameters to evaluate facility investments and the effectiveness of new technologies.

It was stated previously that the proposed model should be able to identify the critical links with excessive queuing (delays) due to either recurrent congestion or incident conditions within a traffic
On this basis, the model should be able to divert some traffic from their original routes to alternative routes in order to reduce the overall network delay. As stated previously, the diversion process can be either by a one-time period assignment or by a multiple time period approach. Under the multiple time period approach, the explicit network link queuing information from each equilibrium solution of the model should be considered in the next iteration, because the leftover queues within the network have significant impacts in the new time increment on both the travel times on alternative routes and on the overall length of time it takes for the origin-destination demand to be satisfied.

11.4.0 RESULTS OF THE REVIEW

The initial evaluation of each software package in the review list was conducted in two stages. In the first stage of the review, the features of each package were summarized in tabular form. Then, in the second stage, a concise evaluation summary was prepared for each package considered.

The results of the initial evaluation are summarized in Table 2. The review summary of each package appears in the following paragraphs.
* Any equilibrium assignment can estimate a system optimum flow pattern if link cost functions represent marginal costs. However, if this is done, the network performance measures representing travel times and delays would be calculated incorrectly.
11.4.1 Summary Description of Individual Packages

11.4.1.1 MINUTP

This package was developed for the purpose of comprehensive urban transportation planning. It is capable of performing both highway and transit network modelling for transportation studies. The package is designed with a very large capacity to handle large networks. However, this package does not have the capability to perform detailed arterial and intersection level of service modelling tasks.

With reference to the features desired in the proposed model, MINUTP has the ability to perform equilibrium trip assignment. The average network performance parameters can be estimated. These estimated values are confined to the major highway and arterial network only, since the package does not have the capability to model the surface street network in detail. In its unmodified form, MINUTP does not have the capability to analyze explicit queue lengths on network links, or to perform multiple time period trip assignment. This restriction to single time period analysis is reflected in the matrix manipulation capabilities of the program. Traffic pre-loading can be accomplished with specific program commands in coordination with the trip assignment module.

An interactive graphics module for network data entry editing and volume display is available. Network data structures are flexible
in the sense that multiple user-defined link variables can be accommodated.

11.4.1.2 TMODEL
This package is also designed for urban transportation planning applications. It has the capability for highway network modelling and surface arterial network modelling, but the program capacity is relatively small and can not accommodate large networks. The program does not have the capability to perform equilibrium trip assignments. All-or-nothing trip assignment is available, and assignments can be conducted under capacity restraint with percent increments. Network link queue length analysis is not available. It should be especially noted that the program has interactive dynamic intersection capacity analysis capabilities. The surface street intersection database can be downloaded to intersection capacity analysis programs. The analysis results can provide detailed information on intersection performance.

The latest version of this software package has the capability for interactive on-screen graphics network editing and data entry.

11.4.1.3 TPANPLAN
TRANPLAN is an urban transportation planning package with very large capacity to handle both highway traffic network modelling and transit network modelling. Like MINUTP, it does not have the
capability to perform detailed surface street traffic network modelling.

TRANPLAN'S equilibrium trip assignment method is identical to the method employed in MINUTP, and it has very similar output of the network performance parameters determined through running trip assignment. There is direct pre-loading of base traffic available. Again, link queuing analyses is not available. The capabilities of selected links and subsection analyses are available. An interactive graphics network editor is provided in the latest version.

11.4.1.4 CARS
This package was developed for metropolitan traffic and land development impact modelling with a relatively small network capacity. The significance of this package in comparison with the others in the review list is that it has the capability to pre-load traffic flows to all nodes with explicit turning movements. Then, after user-defined alternative paths or minimum path trip assignment, the node traffic volumes can be downloaded to another interactive computer program for detailed intersection capacity analysis. CARS offers very limited information on operating conditions on freeway links.
Besides its special features of dynamic intersection modelling, it does not offer as many network modelling capabilities as most other packages compared in the review.

11.4.1.5 MicroTRIPS

This package is another comprehensive computer software system designed for transportation planning. It has functional capabilities parallel to the UTPS system. Its network modelling capacity varies with the capacity of the computer used to run the system.

Similar to MINUTP, capacity restraint trip assignment can be applied to network modelling. However, the current distribution offers no equilibrium assignment. A flexible network traffic pre-load function is available. In addition, the new trips assigned to a preloaded network may follow any available assignment principle, which may differ from the principle employed for preloading the network. Subsection analysis and interactive graphics are available.

However, the package does not have the capability to analyze the surface street system in a detailed manner. According to the software distributor, an interactive dynamic intersection analysis package is under development for integration with the current programs. In addition, a dynamic network assignment procedure, incorporating multiple time periods and explicit queuing, is
reported to be available, although this capability was not included in the version reviewed. Further investigation of this new capability is warranted, although indications to date suggest that the microTRIPS Dynamic Assignment is not based on the equilibrium principle.

11.4.1.6 EMME

This software package was developed for urban transportation planning studies with emphasis on urban streets and transit network modelling. It is equipped with more powerful matrix manipulation capabilities than other packages. Its highway trip assignment method is limited to equilibrium assignment. Subsection analysis can be performed through the matrix manipulation module, but selected link analysis is not directly available.

The strength of the EMME package is its fully interactive on-screen graphics-based operation and its unique transit assignment method. Surface street intersection capacity analysis is not available in this package.

11.4.1.7 FREQ

The FREQ program was developed for freeway corridor traffic flow simulation, optimization, and control. It is specifically designed to develop system-optimum control strategies within the freeway corridor environment. It has the capability to analyze freeway corridor flows and explicit link queuing. In addition, the excess
demand from link queuing can be carried over to the next time increment of trip assignment within the corridor.

However, this program only can accommodate one freeway and a parallel arterial in the network model. It is not able to perform network modelling. An alternative route can be identified between the freeway and a parallel arterial, but the detailed operating parameters for the arterial can not be assessed. An interactive graphics program for data entry is available in the latest version.

11.4.1.8 NETSIM

This program was developed for local area traffic network operational analysis. It was mainly designed for surface street networks which is just the opposite of the freeway corridor-oriented FREQ package.

This network simulation package has the capability to analyze surface street network operations, especially intersection operating conditions, in a microscopic manner. Explicit queuing analysis can be performed at each intersection and on connected links. The package offers various analysis capabilities based on user-defined link flows. It does not have a route choice capability in the program.
11.5.0 CHAPTER SUMMARY AND CONCLUSION

Through an initial evaluation of computer software packages, some general conclusions can be reached for the next step in our model development process.

Four of the transportation planning-oriented packages reviewed have equilibrium trip assignment capability with various useful options, such as network preloading, selected link analysis, and percent incremental loading. All six packages have the capability to estimate network flow parameters to various degrees. However, none of them are able to determine explicit queue lengths on network links, nor deal conveniently with demand carry over in a multiple time period assignment.

Two packages, TMODEL and CARS, are developed with interactive computer programs to assess surface street intersection operating efficiency. However, neither package has the traffic pre-load option.

Interactive on-screen network editing is available in five packages. This feature lets users make direct visual examination of the coded network and edit the network.

The two traffic operations analysis-oriented packages have their own special characteristics. FREQ emphasizes freeway corridor simulation and system optimization, and NETSIM is designed to
simulate surface street traffic operating within a microscopic environment. Both programs can provide detailed information on corridor or network performance, although neither incorporates any route selection capability. Interactive graphics is not available with either.

In conclusion, considered as a group, the programs which were reviewed generally possess most of the features desired in the proposed model. However, each one lacks certain critically important features. The network planning packages provide equilibrium assignment but lack multiple time period analysis and explicit queuing, and they deal awkwardly with system-optimum assignments. The FREQ and NETSIM models consider the dynamic aspects of traffic flow and explicit queuing, but have no network-wide route choice capability. None of the packages considered appears able to produce a constrained system optimum assignment, which is important to the objectives of this research.

The principal conclusion is that it appears necessary to develop, more or less from scratch, a network assignment procedure with the desired capabilities. An alternative would be to expand FREQ to become a network analysis model, incorporating the type of multiple time period, dynamic equilibrium assignment procedure described in the following chapters. At this stage, it is not clear whether there would be any savings of effort achieved by building upon the existing FREQ program. Whether it is decided to expand FREQ or to
develop an entirely new program from scratch, it does seem desirable to use network and matrix data structures compatible with one of the existing planning packages, like MINUTP, in order to take advantage of the interactive graphics and additional modelling features already available in these packages. Based on past experience, implementing MINUTP-compatibility appears to be a suitable option.
III: LINK COST FUNCTIONS INCORPORATING EXPLICIT QUEUING

111.1.0 INTRODUCTION
The equilibrium trip assignment methods found in many network analysis software packages require the specification of link cost functions, also called travel time or impedance functions. These functions appear as shown in Figure 1. They attempt to show how link travel time varies with increased flow, usually measured in vehicles per hour. The portion of the function where demand flow, \( D \), is less than saturation capacity, \( C \), is equivalent to the top part of the standard speed-volume curve for an uninterrupted flow facility, like a freeway. The rest of the function, where \( D > C \), attempts to represent the bottom part of the speed-volume curve, by capturing the extra delay experienced by traffic during forced flow, or queuing conditions.

However, this approach to modelling queue delays totally ignores the dynamic nature of the problem. Specifically, during a certain steady state time interval when demand, \( D \), exceeds capacity, \( C \), the average delay due to time spent queuing depends both on the duration of the time interval and the length of the queue, if any, when the time period began. Essentially, for any time period during which demand to cross a link is assumed constant, the cost function is in reality undetermined, since it depends on the queuing history of the preceding time period. This is also true
of the part of the function where \( D < C \), if a queue exists at the beginning of the time period under consideration.

In general, the dynamic aspects of queuing should not be ignored in transportation planning modelling, because the excess demand on certain links may disrupt upstream flow and cause extremely large delays. Also, to be able to analyze alternative strategies to alleviate queuing problems, it is necessary to deal with the incidence and timing of oversaturated flow conditions within a network in some detail. To do this properly, it seems necessary to be able to use an equilibrium trip assignment procedure on an incremental time-slice basis. With this approach, the locations and approximate timing of link queuing relative to different time slices can be determined. Currently available transportation network planning computer packages do not have the capability to perform this type of dynamic equilibrium assignment modelling.

A purpose of this report is to specify a mathematical algorithm suitable for computer implementation which will be able to deal in a fairly realistic manner with the excess demand problem on network links. On the other hand, it is important that the algorithm and network representation be simple enough that data requirements are appropriate for planning applications, and computation time is tolerable for fairly large networks (on the order of one to two hundred centroids, and one to two thousand links).
This chapter presents an approach to formulating link cost functions within the framework of a multiple time period equilibrium assignment model. The assignment procedure itself is described in the chapter that follows.

In succeeding sections, following a statement of assumptions, a description is provided for the basic form of the link cost function and its associated variables. Derivation of the generalized link cost function, which explicitly considers the average delays due to queuing, is presented in stages, by considering four different combinations of demand and initial queue conditions for the time increment being considered.

111.2.0 ASSUMPTIONS

The following assumptions are considered necessary in order to develop dynamic link cost functions appropriate for a planning model:

1. Traffic flow is considered uniformly distributed during each time increment.

2. The saturation flow rate of network links is constant and equivalent to the link saturation capacity.

3. The average link travel time at the saturation flow rate is constant.
4. Each network link is represented by its unique link cost function, although a small number of generalized functional forms can be applied to the entire network.

5. Although, in reality, a queue forms at the entrance of any link for which demand exceeds link capacity, and may extend a considerable distance upstream, for planning application, each queue is assumed to be fully contained at the entrance of the bottleneck link, and no spillover effects on the performance of upstream links are considered.

6. Flows on links in each time increment are solely determined by the demand matrix for the time increment and the proportions of the O-D paths which can be completed in the time available during the increment. Link flows are not directly constrained by upstream bottlenecks except to the degree that the bottlenecks increase path times.

111.3.0 DERIVATION OF THE DYNAMIC LINK COST FUNCTION

Each link cost function represents the relationship between the link's average travel time, its demand, and its capacity. The generalized form of the function is usually presented as:

1.a \[ T = T(D, C); \]

\[ T: \text{link travel time} \]
\[ D: \text{link demand} \]
\[ C: \text{link saturation capacity} \]
The standard link cost function can be graphically illustrated as shown in Figure 1.

**FIGURE 1: Link Capacity Function**

As can be observed in the figure, the link travel time increases with an increase of link demand.

As long as demand remains less than saturation capacity, shown in Figure 1 as $C$, travel time remains fairly constant, increasing only slightly as demand approaches capacity, above a threshold $C_p$ which at one time was inappropriately termed "practical capacity." As long as demand remains less than $C$ and no queues are present, the portion of the link cost function where $D < C$ is a reasonable representation of the average conditions which would be expected to occur on an actual roadway section.
There are several ways to look at the part of the link cost function where \( D > C \). The harshest way is to conclude that this portion of the curve is meaningless for steady state conditions expected during short time increments, since link flow simply cannot exceed capacity. The conventional approach, as described by Branston (23), acknowledges that planning models do not really deal with steady state conditions, but rather consider rather lengthy time periods, during which the demand rate of flow to enter each link fluctuates, first increasing, then decreasing until any queue which may have formed during the period has disappeared. From this basis, the demand rate matches some short peak time interval during the overall time period to be modeled, and the right-hand portion of the function attempts to represent the extra delay involved in queuing, both during and subsequent to the peak interval, as the queue eventually disappears.

The problem with the conventional approach, as indicated previously, is that the amount of queue delay is more than a function of demand and saturation capacity. It also depends strongly on the time distribution of demand during the overall time period to be modelled. By implication, average link time also depends on the length of the time period being considered. Neither of these variables is considered in the conventional static formulation shown previously.

The alternative formulation which we propose is the following:
lb. \( T = T(D, C, t, D_i) \); 
\( T \): link travel time 
\( D \): link demand 
\( C \): link saturation capacity 
\( t \): length of the time period in which \( D \) is experienced 

\( D_s \): the excess demand corresponding to the queue present on the link at the beginning of the time period.

Procedurally, it is expected that the entire time period considered in the traffic assignment, \( \tau \), is divided into "\( m \)" equal length time intervals:

2. \( t = \tau / m \)

Generally, \( \tau \) would be from one to three or more hours, depending on the duration of the peak traffic period. The time increment, \( t \), would be relatively short, perhaps as short as five or ten minutes and probably never more than thirty minutes. The key issue is whether \( t \) is short enough that conditions within the increment can reasonably be considered steady state.

Assume that an equilibrium traffic assignment can be performed to establish routes for each time increment and that, through some suitable accounting process, an appropriate estimate can be made of the demand rate to enter each link (in vehicles per hour). In general, that demand is comprised of some portion of the trips that actually begin their journeys during the current time increment,
and some portion of the trips left over from previous time increments which did not progress to the link in question until the current time increment. The equilibrium assignment and demand accounting procedures required to accomplish this are described in more detail in the next chapter.

The question now at hand is the specific functional form of the link cost function used in this dynamic equilibrium assignment framework. Clearly, any such function must begin with the base cost function which represents the steady state relationship shown in Equation 1a for $D < C$ and $D_e = 0$ (no initial queue). Such a base cost function is strictly empirical, depending on link length, geometries and other driving conditions. In the discussion which follows, we will use $T_b(D, C)$ to represent the empirical value of the base cost function, which is completely equivalent to the top part of the standard empirical speed-volume relationship.

In order to develop appropriate expressions for the dynamic link cost function of Equation 1b, four steady state conditions are examined:

I. $D_e = 0$ (no initial queue) and $D > C$

II. $D_e > 0$ and $D > C$

III. $D_e > 0$, $D < C$, and any initial queue is too long to disappear during the time increment $t$
IV. $D_e > 0$, $D < C$, and any initial queue does disappear during the time increment $t$.

There is, of course, one additional case, where $D < C$ and $D_e = 0$. This is simply the base link cost function described previously.

111.3.1 Case I: No Initial Queue, But $D > C$

In Case I, it is assumed that the base link cost function is as illustrated in Figure 1 under unsaturated flow conditions. The near free flow link travel time is $T_p$ and the link saturation flow travel time is $T_s$. Under Case I, there is no link queue at the beginning of the time increment, but the link demand rate, $D$, during the increment is greater than the saturation capacity $C$. During this time increment, the total number of vehicles in the demand, $F_d$, and the total number of vehicles in the actual flow allowed by the link capacity, $F_c$, can be determined as:

5. $F_d = D \times t$, and
6. $F_c = C \times t$

Based on these parameters, a link queuing diagram can be established for the assignment period as shown in Figure 2.
As shown in the figure, when the link demand exceeds its capacity, a queue is gradually formed. At the end of the time increment, a queue length denoted as $Q$ (vehicles) is present.

7. $Q = F_d - F_c = (D - C) \times t$

This queue length gives rise to the notion of "link excess demand" for the time increment. Let the link excess demand be denoted as $D_e$ in vehicles per hour, then:

8. $D_e = \frac{Q}{t} = D - C$

Graphically it can be observed that the vehicles in the queue eventually traverse the link, usually at the beginning of the next time increment. They therefore affect the link cost function for the next time increment. Consequently, the value of link excess
demand, \( D_{t} \), used in the link cost function for a particular time increment is that which is calculated based on conditions at the end of the preceding time increment.

Using the queuing diagram, the link maximum delay (MD) and average queue delay (AD) are geometrically determined in terms of the link demand, capacity, and the duration of the time increment:

9. \[ MD_{t} = [(D / C) - 1] * t \quad ; \quad D > C \]

10. \[ AD_{t} = (1/2) [(D / C) - 1] * t \quad ; \quad D > C \]

As can be seen in the diagram, the maximum delay is experienced by the last arrival during the time increment. The average queue delay is used as the additional perceived link travel time in excess of the base link travel time at the saturation capacity, \( T_{s} \). Then the link average travel time, \( T_{t} \), in the oversaturated condition can be expressed as:

11. \[ T_{t} = T_{s} + AD_{t} = T_{s} + (t/2)[(D/C) - 1] \]

Since the link capacity function where \( D > C \) is defined as an extension of \( T_{s} \), it can be represented by a linear extrapolation of the base cost function with a constant slope of \( (t / 2C) \) as determined from the link average delay function. This link capacity function is illustrated in Figure 3.
By utilizing this link capacity function, the average link travel time can be identified under oversaturated conditions, provided that no queue exists at the beginning of the time increment in question.

111.3.2 Case II: With Initial Queue and D > C
Under this condition, the link demand is not only the demand of the current time increment, but also the excess demand from the previous period. The link queuing diagram for this condition is illustrated in Figure 4.
As shown in the diagram, the initial queue, $Q_o$, which comes from the link excess demand of the previous time increment, has to be cleared before the current demand, $D$, can be served by the link. Therefore, the link total demand $TD$ is the sum of the current demand and the previous link excess demand $D_e^0$.

12. $TD = D + \left( \frac{Q_o}{t} \right) = D + D_e^0$

Similar to Case I, at the end of this time increment, a link queue, $Q$, also exists. Its length can be determined from the diagram as:

13. $Q = (D - C) \times t + Q_o = (D + D_e - C) \times t$

Similar to Case I, the link excess demand problem can be assessed in terms of the link delays.

14. $MD_e = \left\{ \left[ \frac{(D + D_e^0)}{C} \right] - 1 \right\} \times t$, and

15. $AD_e = (1/2)\left\{ \left[ \frac{(D + 2D_e^0)}{C} \right] - 1 \right\} \times t$
Using the same approach as in Case I, it is clear that the average travel time under oversaturated conditions when a queue exists at the beginning of the time increment can be expressed as:

16. \[ T_2 = T_s + AD_s = T_s + \left( \frac{t}{2} \right) \left\{ \frac{(D + 2D_o)}{C} \right\} - 1 \]

Note that this expression is equivalent to Equation 11 for \( D_o = 0 \).

111.3.3 Case III: With Long Initial Queue and \( D < C \)

Under this case, the analysis of the queuing problem is very similar to that of Case II, except that the initial queue shrinks during the time increment, because the link demand is less than the link capacity. The queuing diagram for this condition is illustrated in Figure 5. Although the link arrival rate is less than the link capacity, the link still has a queue at the end of the increment because the total link demand in the increment is still greater than the capacity.

The maximum link delay is geometrically determined from the diagram as:

17. \[ MD = \left[ \frac{D_o}{C} \right] * t \]

It can be determined from the diagram that both the average queue delay and the average travel time for this case are the same as for Case II, expressed by Equations 15 and 16.
111.3.4. Case IV: With Short Initial Queue and $D < C$

The significant difference in this case from the previous one is that the total link demand within the time increment is less than the link capacity. This can happen because the initial queue is small, because $D << C$, or from a combination of the two. Whether the queue delay experienced by part of the demand at the beginning of the increment is substantial depends on the initial conditions. This phenomenon can be observed from the queuing diagram in Figure 6.
As shown in the diagram, the link departure rate is maintained at its maximum to accommodate the initial queue from the previous time increment and the additional demand of this time increment until the queue has disappeared. Then the link departure rate becomes the same as the link arrival rate, through the end of the time increment. During the portion of the time increment when the queue exists, the link travel time (ignoring queue delay) is $T_s$; during the rest of the increment, the travel time is given by the base link cost function, $T_b(D, C)$.

According to the queuing diagram, the maximum delay and average queue delay are calculated as:

18. $MD_d = \left[ \frac{D_e^o}{C} \right] \ast t$, and

19. $AD_d = \frac{1}{2} \left( \frac{(D_e^o)^2}{[(C-D)(D+D_e^o)]} \right) \ast t$

The average link travel time is therefore the sum of the average queue time and the non-queue travel time, which is a weighted average of $T_s$ and $T_b(D, C)$, where the weights are the number of vehicles crossing the link (1) when the queue exists and (2) after the queue has disappeared. The amount of time before the queue disappears is calculated as:

20. $t_q = \frac{D_e^o \ast t}{(C - D)}$
This leads to the following expression for the average link travel time:

\[
T_4 = \frac{D_e^o + 2T_s + C*D_e^o + 2T_b(D,C)}{2(C - D)(D + D_e^o)}
\]

Neither the expression for average queue delay nor the expression for average total link time are monotonically increasing functions of D. Rather, they are U-shaped. This presents a serious problem because, if these link cost functions are used directly, they violate the conditions required to force the equilibrium assignment to reach a unique solution.

The question arises whether it is possible to develop an approximation to the actual link cost function in Equation 21 which has the property of being monotonically increasing, with the same general shape illustrated in Figure 1. It turns out that this is indeed possible. In developing such a function, it is important to recognize certain limiting conditions, specifically, that as \(D_e^o\) approaches zero, \(T_4\) approaches \(T_b(D,C)\), the standard static link cost function illustrated in Figure 1. At the other end, as \(D_e^o\) approaches link capacity, \(D\) approaches 0 and \(T_4\) approaches the constant value \(T_s + (t/2)\). In between these extremes lies a family of U-shaped curves, generally with a short decreasing portion followed by a lengthy increasing portion similar in shape to \(T_b(D,C)\), although much higher valued. As \(D_e^o\) increases, the curves
rise and flatten to become nearly constant-valued, asymptotic to $T_e = (t/2)$. Appendix 1 contains several graphs showing the nature of this family of curves for different values of $D_e^o$, expressed as a percentage of the capacity, C.

The appendix also shows the shape of an empirical function $T_4'$ which provides a close approximation to $T_4$ across all permissible values of its parameters. The expression for this best-fit function is:

22. $T_4' = T'(D, C-D_e^o) + [T_4(C-D_e^o, C, t, D_e^o) - T'(D, C-D_e^o)] \times \text{LOGIT}$

where:

$$T'(D, C-D_e^o) = T(0, C-D_e^o) + \frac{T(D, C-D_e^o) - T(0, C-D_e^o)}{T(C-D_e^o, C-D_e^o) - T(0, C-D_e^o)} [T_4(C-D_e^o, C, t, D_e^o) - T(0, C-D_e^o)]$$

(The $T'(\cdot)$ function is the base link cost function $T(\cdot, \cdot)$ scaled vertically to reach the same maximum value as the $T_4'(\cdot)$ function evaluated at $D = C-D_e^o$.)

$$\text{LOGIT} = \frac{4 \times D_e^o/C}{4 \times D_e^o/C} - 1 \quad 4 \times D_e^o/C$$

(The purpose of the LOGIT term is to cause the $T_4'(\cdot)$ function to approach a constant value equal to the maximum value of the $T_4(\cdot)$ function, as $D_e^o$ approaches capacity.)
It is worth noting that when $D_e^o = 0$, $T'_4()$ is equal to the base cost function, Equation 1a. Furthermore, when $D + D_e^o = C$, both $T_4()$ and $T'_4()$ equal the value of the cost function for Cases II and III, calculated by Equation 16.

111.3.5 Summary of the Cost Function Analysis

Four possible cases of link queuing have been analyzed, and exact solutions obtained for all. A close approximate solution which provides a monotonically increasing functional form was developed for the one situation (Case IV) where the exact solution is not monotonically increasing.

It turns out that the solutions for two of the queuing cases, as well as the static base cost function itself (Equation 1a) are special cases of the others. Consequently, the full set of solutions reduce to a single continuous monotonically increasing link cost function, suitable for use in a dynamic equilibrium assignment algorithm. The final link cost function, $T(D, C, t, D_e)$, defined for the entire domain of $D > 0$, is the following:

Where $D + D_e \geq C$ (a queue exists at the end of the time increment):

Use Equation 16;

Otherwise:

Use Equation 22.
Based on the considerations of preceding chapters, a detailed modeling framework is now presented. This chapter deals with the data structure needed to perform the analysis, as well as data processing and algorithmic aspects of the proposed model. It sets out many aspects of the work to be accomplished in order to implement the analysis capabilities described in this report.

**IV.1.0 DATABASE**

It is necessary to choose an appropriate database structure for storing the required network and traffic flow information, and for transferring necessary information to other computer programs. The database will store data with the link as the basic unit. Since a given network may contain freeways and arterials, and signalized intersections on arterials, the database will conform to the different facilities' characteristics in their data structure. The characteristics would include, but not necessarily be limited to the following (both inputs and outputs of the assignment are listed):

Freeway: link volumes (by direction; by time increment)
link length
number of lanes
grade
queue length (upstream of the link)
Arterial: link volumes (by direction; by time increment)
  link length
  number of lanes
  peak hour factor (PHF)
  arterial class (by HCM classification)
  queue length (before the first signalized intersection within the link)
  signalized intersection characteristics at the end of the link:
    cycle length and phasing
    saturation flow rate per lane
    number of lanes per turning movement
  progression adjustment factor (per HCM)

Using these data, it should be possible to estimate fairly accurate capacity values and speed-volume curves for each link to be represented in the urban network. In principle, these data should be able to be accommodated in the MINUTP network data structure, although the numerous extra variables required for arterials suggests that some of these data may more efficiently be stored elsewhere.

In order to perform the trip assignment, a peak period trip matrix file will be needed within the database. For multiple time period analysis, a time-of-day distribution over trip starting times will be required, most likely varying for different subsets of origin
zones. Again, the basic matrix data structure should follow the MINUTP convention.

IV.2.0 NETWORK GENERATOR
In order to convert the information in the database to capacity and speed-volume curves (the base link cost function T(D,C)) to be used in the network model, several capacity and level of service analysis programs will be required. These programs, based on established freeway, arterial, and intersection analysis procedures (generally following 1985 Highway Capacity Manual (HCM) methods), should be integrated with the database and a network-generation executive program. This integrated program will retrieve the data from the database to establish the necessary network operating characteristics for each assignment.

The network generator will be used to create the needed capacity values and base cost functions at the start of each assignment. For multiple time period assignments, the link cost function for each time increment will be automatically generated, based on the upstream queue length at the end of the preceding time increment.

IV.3.0 EXECUTIVE PROGRAM
The Executive Program will retrieve the network data from the network generator and perform the desired trip assignment. The Executive Program will be able to use the basic dynamic equilibrium assignment algorithm in a variety of ways, in order to implement
the several network assignment approaches which have been identified. These are:

a.1 Criteria: 1. user-optimum
2. system-optimum
3. constrained system-optimum (to limit deviation from the user-optimum)

b.1 Methods: 1. single time increment (static) assignment
2. multipletimeincrement (dynamic) assignment

Both methods can be applied under all of the different criteria to test the sensitivity of the results to the characteristics of the assignment.

After determining the assignment of traffic in the network, a re-evaluation of the network performance may be needed. This would reconcile certain assumptions involved in estimating capacities and the base link cost functions to the flows and turning movements actually observed in the network.

IV.4.0 THE DYNAMIC EQUILIBRIUM ASSIGNMENT ALGORITHM
The dynamic equilibrium assignment method is a straight-forward extension of the standard equilibrium assignment algorithm found in many transportation planning packages, like UTPS, MINUTP, EMME2, and TRANPLAN. The equilibrium-seeking logic is the same; the differences are that (1) the proposed method reduces each shortest
path to the length which can be reached during the time increment under consideration, subject to an assumed uniform distribution of start times during the increment, and (2) traffic is loaded only on the links of each shortest path that fall within the portion of the path which can be reached during the time increment. In addition, after the equilibrium assignment procedure is applied for each time increment, a new pseudo-O-D matrix is generated for the next time increment, containing the fraction of peak period trips which begin during that time increment plus the remaining portions of all trips which began but were not completed during preceding time increments. That pseudo-O-D matrix, and the link cost functions established according to the method described in Chapter III, are the inputs to each iteration of the equilibrium assignment procedure for a given time increment.

The following variables and functions are used in the algorithm:

- $C_l$: The capacity of link 1, usually expressed in vehicles per hour.
- $D_l^o$: The original link flow on link "1", also in vehicles per hour.
- $D_l'$: A newly calculated link flow on link "1",
**D**<sub>el</sub>  The link excess demand from the previous time increment, defined by Equation 8, in Chapter III.

**T**(D)  The link travel time (or cost) expressed as a function of the flow value \( D \). As noted in Chapter III, \( T(D) \) is also a function of the link capacity, the length of the time increment, and the queue length at the end of the preceding time increment, but these additional parameters are omitted because they are constant for the time increment under consideration.

**t**  The length of the time increment.

**α**  A parameter whose value is obtained during the course of equilibrium assignment, used to establish a new flow value for each iteration.

**K**  A convergence criterion for the equilibrium assignment procedure.

There exist four types of nodes in the network:

0 = 1, 2, ..., 0  Origin centroids, where trips begin.
e = 1, 2, ..., E Destination centroids, where trips end.
(Usually, the origin and destination centroids are the same locations, but they need not be.)

h = 1, 2, ..., H "Holdover" pseudo-centroids, which are nodes selected to serve as intermediate locations along the routes connecting origins and destinations, where long trips resume their journeys at the beginning of a new time increment. We would expect that there would be 3-4 times as many h nodes as origins and destinations. h nodes also usually represent physical features like intersections and locations of link capacity changes.

n = 1, 2, ..., N Other nodes, which represent only physical features like intersections and locations of link capacity changes.

This leads to the definition of a two-part trip matrix for each time increment, as follows:

A An OxE matrix containing the number of trips which begin during the current time increment at each node o and are destined to each destination node e. (This matrix is calculated for each time increment by multiplying each
row of the overall origin-destination matrix by the proportion of trips during the current time increment.)

\[ B \] An HxE matrix containing the number of holdover trips from the previous time increment which originate at each pseudo-centroid \( h \) and are destined to destination node \( e \).

\[ B' \] Another HxE matrix containing the number of holdover trips from the current time increment.

\[ B_w \] A temporary working version of matrix \( B' \), used to accumulate trips within the equilibrium assignment algorithm.

To understand how matrix \( B \) is calculated by the equilibrium assignment algorithm, for use in the next time increment, consider a single cell of matrix \( A \), containing the number of trips, \( a_{o,e} \). Within the algorithm, a shortest path is calculated between nodes \( o \) and \( e \), and several pseudo-centroids, which we will call \( h = 1,2,3, \) are identified along that path. (Note that, unlike regular centroids, flow can pass through pseudo-centroids.) Assume, for this illustration, that the length of the path is less than \( t \), the length of the time increment. On this basis, we define:
$c_h$ The cumulative time along the path from origin $o$ to pseudo-centroid $h$.

$c_e$ The cumulative time along the path from origin $o$ to destination $e$.

$f_h$ The fraction of $a_{o,e}$ that progresses as far as pseudo-centroid $h$ during the current time increment.

$f_e$ The fraction of $a_{o,e}$ that progresses all the way to destination node $e$ during the current time increment.

Given that trips are assumed to depart node $o$ according to a uniform departure distribution, we develop a solution for the $f_h$ values in a manner illustrated by the distribution shown in the following figure.

This figure illustrates that the fraction of trips that begin before time $t-c$, progress all the way to the destination, while
trips that begin after that time get only as far as one of the pseudo-centroids. Note that all trips are assumed for computational convenience to get at least as far as the first pseudo-centroid, a distortion which should not affect the results appreciably. The application of the principle to longer paths should be self-evident.

Knowing the fractions of $a_0e$ stopping at each node $h$, the product $f_ha_0e$ can be accumulated to each cell $b_ne$ of matrix $B$ for the next time increment. (This isn't exactly how it works, the precise method is described within the steps of the algorithm, given below.)

The steps of the assignment algorithm for each time increment are the following:

0. (Initialization) Specify each link cost function according to the principles described in Chapter III, using $D_{el}$ as calculated from the lengths of queues existing at the end of the previous time increment. ($D_{el} = 0$ for all links in the first time increment).

1. Set $D_l' = 0$ for all links, and the array $B_w$ to all zeros. Loop on all origin nodes, $o$. For each, calculate the shortest paths to all destinations, $e$. Then process each $o$-$e$ pair as follows.
a. Using the uniform departure time assumption, determine the fractions of trips, $a_{oe}$, which finish the time interval at each pseudo-centroid along the path, calculated as described above. These are the variables $f_h$.

b. Accumulate the trips to each pseudo-centroid along the path as follows:

$$B_w = B_w + a_{oe} * f_n$$

c. Accumulate the link flows in accordance with the same fractions as follows.

$$D'_l = D'_l + a_{oe} * (1 - \sum_{k \in \Omega_l} f_k)$$

where $\Omega_l$ is the set of pseudo-centroids between link "1" and origin node $o$, including any pseudo-centroid at the beginning of link "1".

2. Loop on all pseudo-centroid nodes, $h$. For each, calculate the shortest paths to all destinations, $e$. Then process each $h$-$e$ pair as follows.

a. If the total path length to destination $e$ is less than the length of the time increment, $t$, accumulate traffic
on all links of the path between node h and destination e as:

\[ D_i' = D_i' + b_{he} \]

where \( b_{he} \) is the cell value from matrix B.

b. If the total path length to e is greater than \( t \), determine the last pseudo-centroid, \( h' \), along the path which can be reached within time \( t \). Accumulate trips to that pseudo-centroid as:

\[ B_w = B_w + b_{he} \]

Then accumulate the link flows on all links between nodes h and \( h' \) as follows.

\[ D_i' = D_i' + b_{he} \]

3. If this is the first iteration through the equilibrium assignment algorithm for this time increment, set \( D_i'' = D_i' \) for all links, and \( B' = B_w \). Then return to step 1. Otherwise, continue with step 4.

4. Through a line-search method, determine the value of \( \alpha \) which minimizes the following objective function:
\[ Z' = \sum_{l} \sum_{l} T(x) \text{d}x \]

where:

\[ 0 \leq \alpha \leq 1 \]

The rationale for this objective function can be found in any text describing the equilibrium assignment method, such as Stopher and Meyburg (21).

5. Set:

\[ D_{i}^{\circ} = D_{i}^{\circ} + \alpha (D_{l}'-D_{i}^{\circ}) \text{ for all links} \]
\[ B' = B' + \alpha (B_{w}-B') \]

If this is the first time through for this time increment, set \( Z' = Z' \) and return to step 1. Otherwise, go to step 6.

6. If \( |Z' - Z^{\circ}| < K \) (convergence achieved) or if the maximum allowed number of iterations has been reached for this time increment, continue to step 7. Otherwise, return to step 1.
7. The equilibrium assignment is complete for this time increment. Output the assignment results, $D_i^*$ and related measures of performance. Set $B = B'$. Calculate the excess demand for each link as:

$$D_{el} = \max(D_i^*-c_i, 0)$$

If this is the last time increment, stop. Otherwise, return to step 0 to consider the next time increment.
FIGURE A1: Approximation to Case IV Cost Curve for $D_e = 1\%$ Capacity
FIGURE A2: Approximation to Case IV Cost Curve for $D_e = 5\%$ Capacity
FIGURE A3: Approximation to Case IV Cost Curve for $D_e = 25\%$ Capacity
FIGURE A4: Approximation to Case IV Cost Curve for $D_e = 50\%$ Capacity
FIGURE A5: Approximation to Case IV Cost Curve for $D_e = 75\%$ Capacity
FIGURE A6: Approximation to Case IV Cost Curve for $D_0 = 90\%$ Capacity
FIGURE A7: Approximation to Case IV Cost Curve for $D_e = 99\%$ Capacity
REFERENCES AND BIBLIOGRAPHY


22. Abu-eisheh, Sameer A. and Mannering, Fred L.*Discrete/Continuous Analysis of Commuters’ Route and Departure Time Choices.* In: Transportation Research Record #1138.
