Estimating the Capacity of Transportation Systems:
A Model with Applications to Freight Transportation

A Report

by

Edward K. Morlok
UPS Foundation Professor of Transportation
assisted by
Stephanie P. Riddle
Research Fellow
University of Pennsylvania

Mid-Atlantic Universities Transportation Center

July 1998

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PB2001-100127

1.	2. REPORT DATE July 1998	3. REPORT TYPE AND DATES COVERED Final Report

4. TITLE AND SUBTITLE
Estimating the Capacity of Transportation Systems: A Model with Application to Freight Transportation

5. FUNDING NUMBERS

6. AUTHOR(S)

Edward K. Morlok and Stephanie P. Riddle

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Pennsylvania
Department of Systems Engineering
Philadelphia, PA 19104-6315

8. PERFORMING ORGANIZATION REPORT NUMBER
111-9810

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Transportation Centers Program
U.S. Department of Transportation
Washington, DC 20590

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Additional funding from the UPS Foundation Professorship in Transportation at the University of Pennsylvania

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The issue of adequate capacity in the freight transportation system to accommodate growing cargo volumes is now becoming a major public policy concern. While the problem of extraordinary rail line congestion and service failures in connection with the recent UP-SP merger has galvanized attention to the problem, the issue extends far beyond the rail system to other modes and contexts. In this paper, the concept of *system capacity* (as opposed to link or facility capacity) is developed, and prior literature reviewed for approaches to estimating it. A model which is developed from the most promising of these is presented. This model is multi-modal in concept, and is intended to be applicable to any vehicular freight mode. The model is tested through application to a portion of the rail network, and results assessed for reasonableness and utility. Finally the applicability of the model and approach to various capacity-related questions is described and discussed.

14. SUBJECT TERMS
Network Capacity
Network Performance
Transportation System Capacity

15. NUMBER OF PAGES
31 pp.

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use of thereof.

TABLE OF CONTENTS

ABSTRACT

The issue of adequate capacity in the transportation system to accommodate growing cargo volumes is now becoming a major public policy concern. While the problem of extraordinary rail line congestion and service failures in connection with the recent UP-SP merger has galvanized attention to the problem, the issue extends far beyond the rail system to other modes and contexts. In this research, the concept of *system capacity* (as opposed to link or facility capacity) is developed, and prior literature reviewed for approaches to estimating it. A model which is developed from the most promising of these is presented. This model is multi-modal in concept, and is intended to be applicable to any transportation mode. The model is tested through application to a portion of the intercity freight rail network, and results assessed for reasonableness and utility. Finally the applicability of the model and approach to various capacity-related questions is described and discussed.

ACKNOWLEDGEMENTS

This research was supported by the U.S. Department of Transportation through its University Transportation Centers Program, and by the UPS Foundation Fund for Advanced Transportation Research. It has benefited from numerous discussions with officers of transportation companies and public infrastructure providers in the U.S., Canada, and the European Community. This assistance is gratefully acknowledged but implies no endorsement of the findings.

INTRODUCTION

Recently Linda J. Morgan, the Chair of the Surface Transportation Board (STB), stated:

> The situation in the West has raised the serious question of whether the transportation system as a whole and the rail infrastructure in particular has adequate capacity to handle the surges in the economy that we have seen in recent months."(Kaplan, 1997)

The issue of availability of adequate capacity of the freight transportation system is a serious one. While it has been brought to national attention very forcefully in recent months by the debacle of clogged rail lines in the southwest, after the merger of the Union Pacific and Southern Pacific Railroads, it is in fact a problem that transcends that specific situation. It is an issue with respect to other modes and contexts as well.

The reason is that freight traffic has been growing rapidly for decades. The origins of this growth are in the dramatic changes in industrial production and distribution that have occurred in recent decades. These include globalization, which results in longer supply chains, and hence longer hauls and more freight ton-km (Schipper et al., 1997). Also, the shift toward mass customization and just-in-time production and replenishment policies leads to demand for faster, more reliable or time definite, transportation. Domestic freight moved by all modes combined (measured in ton-km) increased by 65% from 1970 to 1995 (U.S. Department of Transportation, 1997). In the same period international tons grew by 93%. Cargo carried by all freight modes has increased in this period. Rail freight ton-miles grew by 78%; intercity truck ton-miles by 123%, domestic waterborne ton-miles by 32%, and domestic air ton-miles by 285% (Eno Transportation Foundation, Inc., 1996, p.18). Given the dominance or rail and truck in modal shares (about 40% and 33% of intercity ton-miles respectively), the magnitude of the growth is clearly phenomenal.

Similar growth is occurring for person travel as well. It is well known that citizens of the U.S. travel far more per capita than their counterparts in other developed nations—typically about twice the distance that Europeans travel, for example. Domestic auto travel, representing about 85% of all domestic travel, has grown by 99% from 1970 to 1995, while air travel, representing on a person-mile basis most of the remainder of travel, has grown by 268% (Eno Transportation Foundation, Inc., 1996, p.21). During weekday peak periods most urban highway systems suffer considerable congestion, and in a few large cities public transit systems are also overcrowded to the point that passengers can not board vehicles during such periods.

At the same time the physical movement system has not expanded commensurately. For example, rail track-km has declined by 34% since 1970 and highway lane-km have increased by only 43% (Eno Transportation Foundation, Inc., 1998).

While it is clear that the capacity of the transportation system is becoming an increasingly pressing issue, and that most observers of the system would agree that traffic growth is outstripping expansion of capacity, the fact of the matter is that at the present time we have no operational measures of <u>system capacity</u>. Capacity in transportation has traditionally been measured only at individual elements of the system, such as terminals or links (rail lines, road segments, waterway locks, etc.), and these of course do not measure system capacity.

The purpose of this paper is to present the results of research on developing a quantitative, operational way to measure the capacity of a transportation system. This includes developing an operational means or model to estimate capacity from data that either is normally available or could readily be created for actual systems. The model is then applied to an actual freight network, and this provides a basis for discussing the usefulness of the approach in various logistics, policy, planning, and regulatory settings, and identifying directions for refinement and further development of this approach.

The plan of the paper is as follows. First, the concept of system capacity will be clarified and differentiated from the capacity concepts that are now commonly used in transportation. Then prior research on transportation system capacity will be reviewed, as this provides the basis for the specific methods presented and used in the remainder of the paper. Then the model will be presented, applied, and evaluated. Appendices provide details on the models, including equations and definitions.

SYSTEM VS. LINK OR FACILITY CAPACITY

Capacity concepts and measures now in use in transportation primarily measure the capacity of individual elements of the system. The most common of these measure the capacity of links, intersections (or junctions in the case of rail—modal terminology varies), and terminals. Essentially they provide the maximum throughput per unit time, measured in units of vehicles (or cargo) per hour, given the design features of the facility and the mix of vehicle types. Since throughput (volume) affects travel time and other user costs, capacity is usually defined in terms of various levels of service (as in the highway case) or as a function of transit time. Similar relationships have been developed for terminals. These are described in detail in the literature, including textbooks (e.g., Morlok, 1978) and reports which range from summaries covering all modes (e.g., Cambridge Systematics, Inc., 1998) to detailed ones for individual modes (e.g., Transportation Research Board, 1994 and Kahn, 1979). In all cases the capacity depends upon the mix of vehicle types or cargo types using the facility. Clearly such concepts of capacity do not measure the ability of the system to accommodate traffic, but rather relate to only one facility.

The other capacity concept which has been used is that of corridor capacity. Although the concept has existed for some time, this concept has been developed in the recent National Cooperative Highway Research Project (NCHRP) study (Cambridge Systematics, Inc., 1998) to the point of practical tools for multimodal corridor capacity analysis. It essentially measures the maximum traffic volume that a chain of facilities or links in a corridor can accommodate. If there is only a single path in the chain, as on a freeway or single rail line, then the corridor capacity is that of the lowest capacity element of the chain. In some cases there are parallel links or modes, and then the corridor capacity will equal the sum of the capacities of the various paths, accounting of course for any joint use of the same facility. (This is embodied in the well-known max flow-min cut algorithm of mathematical programming to find the maximum flow possible between two non-adjacent nodes in a network.).

While the concept of corridor capacity is certainly useful, and an important extension beyond facility capacity, it too does not capture the capacity of a system. Rather it can only estimate the capacity of the corridor, either the maximum flow between the end points or some weighted total flow between various origin and destination nodes within the corridor.

However, the development of an NCHRP manual for estimation of corridor capacity underscores the fact that transportation planning and engineering professionals recognize that there is a need to measure capacity at levels more comprehensive than just a single facility, and that there is an urgency in developing this capability. Thus there appears to be recognition of this need not only at policy levels, as suggested by the statement of the STB Chair, but also at the level of transportation professionals.

CONCEPTS OF SYSTEM CAPACITY

In developing measures of system capacity, it is useful first to review the basic concepts of capacity of systems in general, and then turn to specific formulations for transportation systems.

Defining System Capacity

There are basically two definitions of the capacity of a physical system which produces a product of goods or services. One is that the capacity is equal to the maximum quantity of output which the system can produce, considering only physical limitations on production. This output level is of course limited by both the system itself and the environment within which it is functioning. For example, the capacity of an electric power plant is limited by both the number and size of generators and the availability of fuel. This concept of capacity is often termed ultimate capacity.

The other basic approach to defining capacity recognizes that the cost may be far too large at the ultimate capacity for such a level of output to be practically or economically attainable. This suggests another basic definition of capacity: the maximum output at which cost does not exceed a maximum acceptable value. This concept is usually termed practical capacity in the engineering literature. The types of costs that are considered will of course depend on the specific application.

Both of these measures of capacity are used in transportation, to estimate the capacity of a link or intersection (node). These applications all begin with the volume-delay relationship for a facility, as shown in Figure 1. The time required to pass through a facility increases with increasing volume, due to interference. The ultimate capacity is the maximum volume that can be accommodated, as shown. Practical capacity depends upon the maximum delay that can be tolerated, which in practice may be set by policy, as in the case of highway levels of service, or schedule requirements as in the case of rail lines. Monetary costs can be used instead of time, and these may include only time (transformed into the equivalent cost to users) or also include other costs such as those incurred by operation of the vehicle, ownership costs (which vary with the time required), etc. In the case where other costs are included, normally the cost of providing the facility is assumed to be constant per vehicle-km., as it is (approximately) in the case of charges for use of roads via either tolls or fuel taxes. If the facility use cost were to follow the average cost of providing the facility (a declining or a U-shaped curve, depending on circumstances), then the user cost curve might decline over at least some of its range, leading to ambiguity in the definition of practical capacity. It is important to note that environmental costs and other external costs can be included as well (as used in determining the environmental capacity of residential streets, for example).

The practical capacity concept has been widely used at the link and node level, and methods for its use are well developed. However, it is not directly applicable to a network, though many of the concepts underlying it can be incorporated in network analysis, as will be discussed later.

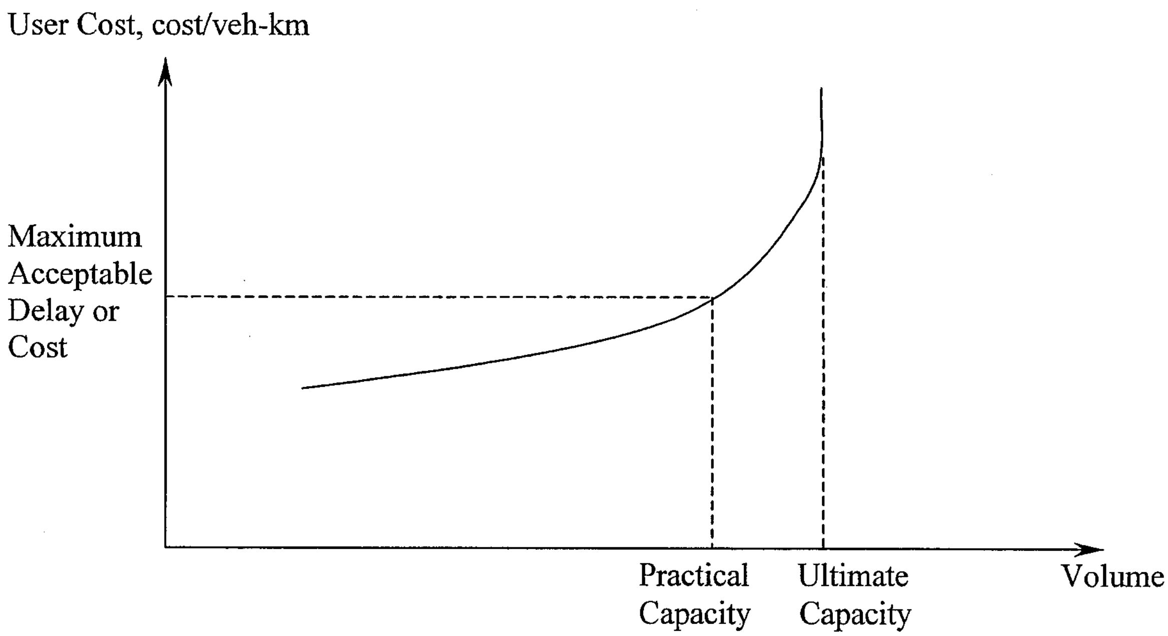

FIGURE 1. Practical and ultimate capacity and the volume delay function.

Another definition of capacity is the <u>economic capacity</u>, defined as that level of output at which the average total cost (ATC) is minimized (DeLeeuw, 1962). Thus capacity is defined in terms of the usual U-shaped ATC curve, as shown in Figure 2. Economic capacity occurs where the marginal cost equals the average cost. Usually this concept is applied to the costs of producer, which in transport would be the costs incurred by the service provider, as opposed to the costs of the service user. Thus this concept starts with different cast basis from the practical capacity notion discussed above. While in principle it could be applied to a link including the same costs, in general the economic capacity notion has been applied using transport carrier or producer costs, vehicle the practical capacity concept has been applied concentrating on user costs.

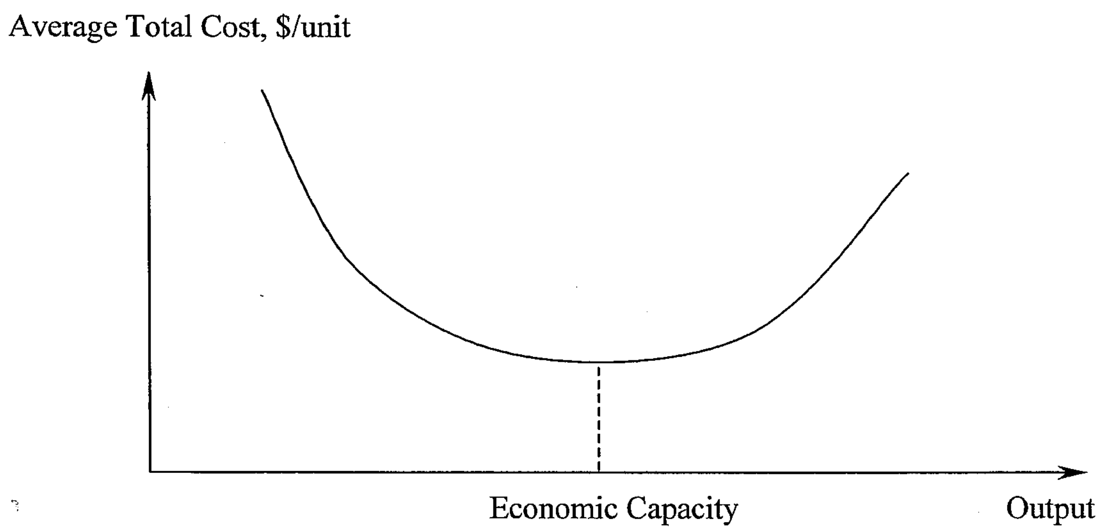

FIGURE 2. Economic capacity and the average total cost curve.

To use the economic capacity definition requires that the cost curve be developed, and in an important paper on economic concepts of capacity, DeLeeuw (1962) points out the potential difficulties in trying to make it operational. From the standpoint of application in transportation, there are two significant difficulties with this approach. First, while it might be applied to an individual firm, it is difficult to see how it could be applied to a system consisting of many firms jointly producing the output of interest, a condition (traffic interchange) characterizing most situations in transport. Secondly, it is defined only for a homogeneous output, so it must be augmented with some specification of product mix if it were applied to a situation like transport. This is because moving a ton of coal is not the same as moving a ton of refrigerators, nor is moving a ton of refrigerators from A to B the same as moving it from B to C. Recognizing difficulties that limit use of the ATC-based capacity definition in practice, DeLeeuw also suggests another definition, as the "point at which the cost of an additional unit of output is well above the most efficient unit cost". (DeLeeuw, 1962, p.834). Thus this variation borrows an idea from the practical capacity approach used in engineering analyses.

Thus there are two distinct lines of thinking about capacity of a production system. One has been developed focusing primarily on user costs, while the other focuses on producer costs. They can be combined, by including the same cost, but in some cases this can lead to internal contradictions, as described above. Also, they are limited and not directly applicable to the network or system level. However, they do provide elements that will prove useful in looking at network capacity.

Estimating System Capacity

The earliest reference to estimating transport system capacity that was found was an attempt in 1972 by Scheppach and Faucett to determine the capacity of the U.S. transport system (Scheppach and Faucett, 1973). This effort found the economic capacity concept problematic for transportation applications at that time. Instead they concentrated on using an empirical measure that they and others have termed normal capacity. This measure "would encompass a number of alternative measures which are based on some past peak utilization or some level of accepted normal operation." (Scheppach and Faucett, 1973, p.4-1). Productivity and cost data, for the period 1950 to 1970, were used to estimate this peak utilization. The observed maximum revenue producing activity per vehicle per year, by vehicle type, was applied to the then-current fleet to estimate capacity of the system in ton-miles per year. Clearly this estimation procedure included only one limitation on traffic carrying capacity, and ignored others such as the facilities infrastructure. Also, it did not deal with the issue of the spatial aspects of flows in the sense of origins and destinations. Numerous caveats qualified the numerical results.

System capacity was also discussed in a transportation economics book by Kneafsey published in 1975 (Kneafsey, 1975), covering the same approaches as presented in the DeLeeuw paper and in the Scheppach and Faucett report. Kneafsey concludes that the concept of normal capacity is flawed, because this highest historical rate may not measure the real maximum output (with or without cost limitations) and also may be based upon inefficient operation by the firm or firms (Kneafsey, 1975, p. 153). He also calls for development of more theoretically sound methods for capacity estimation.

In a 1981 paper proposing a specific quantitative approach to estimating transportation system capacity, Morlok (1981) identifies two problems that must be dealt with. One is that some means must be used to deal with the heterogeneous nature of transportation system output. The second is that some means must be developed to incorporate all of the limitations of various resources on the capacity of the system. These include not only the capacities of facilities but also limited resources in the vehicle fleet, labor, fuel, etc.

To deal with the heterogeneity of output of the system, it is proposed to describe the output by a vector giving the quantity of items transported; by type, and origin-destination city (or node) pair, per unit time. This is consistent with what is done in typical network data bases. Items transported would be either shipments, by commodity, or persons, perhaps differentiated by trip purpose, class of accommodation, or demographic factors, as appropriate. Each category or class usually consists of a collection of similar items, travelling between the same origin area and destination area, each area usually consisting of many towns or other geographic units. It may include reference to the size of shipment, whether of carload or truckload size or smaller, since this determines the transportation processes used and costs, as well as the feasible paths and modes. The precise specification of these categories, and their number, will of course depend upon the particular application. Although terminology is by no means identical in the various sectors of transportation, it has become common to refer to each of these categories as a traffic lane, and this terminology will be used here.

If there are N traffic lanes in the system, then the output of that system could be described by the quantity of traffic in each lane. The output in each category or traffic lane i can be denoted Q_i, measured by weight or cubic volume in the case of freight, persons in the case of passengers transport, or some other appropriate measure. The total system output is thus Q,

$$Q = \sum_{i=1}^{N} Q_i \tag{1}$$

Critical to any specification of overall output is the underlying traffic pattern, λ, which specifies the relative quantity of traffic on each of the traffic lanes (items by class, O-D pairs, etc.). It is a vector, each element of which refers to one traffic lane:

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_i, \ldots, \lambda_N) \tag{2}$$

$$\lambda_i = \frac{Q_i}{Q} \tag{3}$$

The operational procedure suggested by Morlok (1981) for estimating system capacity as defined above is to characterize the problem as a mathematical program, in which a scalar quantity is maximized subject to a number of constraints. In this case, the value of Q is maximized, subject to the constraint on traffic pattern as in eqtn. (3) above. In addition other constraints describe the relevant physical or other characteristics and limitations of the system.

Four types of relationships are identified. The first relationship treats the capacity of facilities to accommodate traffic flowing through them. The second type of relationship deals with the necessity of having vehicles (and containers, etc.) available to accommodate all of the traffic which is to be moved. The third relationship deals with the adequacy of various types of resources that are essentially continuously consumed as the system operates, such as fuel and labor. The fourth relates to costs—either to users as in delays or to system owner/operators-- as a basis for limiting traffic. While the proposed model was not operationalized, it did address the heterogeneity of output issue and also the means of incorporating various constraints and resource limitations in a manner consistent with basic transportation relationships.

A final paper on capacity appeared in 1993, in which Nijkamp et al. (Nijkamp et al, 1993) further developed the idea of many types of constraints that should be considered in any assessment of system capacity. Specifically, they discuss not only the four types of constraints described above, but add environmental regulations and management structure as factors to be considered. While no specific means for operationalizing capacity estimation is presented, by implication it could be incorporated in the mathematical programming framework suggested by Morlok (1981).

THE PROPOSED MODEL

A general model for estimating capacity of a single or multiple mode transportation system can be constructed based on the mathematical structure of optimization suggested by Morlok (1981). The general form of each of the basic relationships will be described below. Each type of relationship will be identified by a shorthand name and symbol to facilitate the presentation of different versions of the model, that address different questions. In this section the relationships will be described, so that the reader obtains an overall understanding of the model. Specific equations, for freight and passenger versions of the model, are presented in an appendix, as the detail and complexity of them can distract from understanding the model. This presentation is followed by an example application.

Model Relationships

Since the terminology for freight and passenger transport differ, the same general types of relationships apply to freight and passenger applications. Here we present the general model, using the term traffic to refer to cargo or passengers. There will naturally be light variations in the model as it is applied to different contexts, such as freight or passengers, and as it is applied to different modes, especially as between vehicular vs. non-vehicular (e.g., pipelines for liquids or conveyors for passengers. The basic similarities, as well as the variations in detail, can be seen from the model equations presented in Appendices A and B.

System Traffic (Type A)

System traffic is measured by total traffic in tons, containerloads, passengers, or a similar unit. When determining system capacity, this is the quantity that is maximized.

Traffic Pattern (Type B)

The traffic flows are always subject to the traffic pattern constraints which ensure that an appropriate amount of traffic is flowing in each of the traffic lanes. The traffic pattern can vary depending upon the specific application and problem. A typical pattern of interest would be the current pattern. Other possibilities will be discussed later.

Arc Vehicle and Traffic Flows (Type C)

These relationships translate cargo or passenger flows into traffic-carrying vehicle flows, referred to as revenue vehicle flows in the person travel context and loaded (not necessarily fully loaded) vehicle flows in the cargo context. Obviously this type of relationship does not appear in the case of non-vehicular modes, such as liquid and gas pipelines, or passenger conveyors. Generally either an average load per vehicle, or a load factor would be used to translate traffic to vehicle flows, and of course the parameter value would vary among different contexts..

Link Congestion and Delay Relationships (Type D)

One set of constraints incorporates the link congestion and delay relationships, which are based on the relationship between the time required for a vehicle to move through or over a facility and the volume of vehicles (or other traffic units) per unit time entering that facility. Such a relationship is often termed a congestion or delay function. This function is related to the design features of the facility, including it size (e.g., number of lanes or tracks), and vehicle movement controls. The exact relationship will naturally vary with the particular mode and type of facility being represented.

Terminal Capacity (Type E)

These are similar in general concept to the arc relationships, in that they represent the limitations on the merging and diverging of traffic and vehicle flows at the terminals. These often result in delays to both traffic and vehicles, and of course there is an upper bound on the volume of traffic and vehicles that can be accommodated. In addition, terminals are usually the location of storage and maintenance facilities for vehicles (and containers), and limitations of such activities would also be incorporated in these relationships.

Fleet Size (Type F)

Sufficient vehicles must be available to accommodate the traffic, or conversely the traffic carried must not exceed the limitations of available fleet operations (the relationship emphasized in early works on capacity, as in (Scheppach and Faucett, 1973). There are two related sets of equations which ensure that this condition is met. One requires that the total time vehicles are available in the period considered (e.g., week, month), measured by vehicle-hours, must be at least as large as the vehicle-hours of vehicle use, whether loaded or empty, and moving or at rest (as at terminals, intermediate nodes, etc.). For loaded movements on each traffic lane the necessary vehicle-hours is determined by the product of total loaded vehicle trips originating at that lane's origin for that lane's destination and the total cycle time required for such movement. The latter is estimated considering the effect of traffic volume on the link and path times incorporated in the facility capacity relationships. Similar relationships exist for the empty movement. In addition, a conservation of vehicle flow equation is needed at each node in the system, specifying that the total number of vehicles of each type entering the node (loaded or empty), must equal the total number of vehicles leaving that node (loaded or empty).

In addition, in some transportation systems such as railroads and inland waterway (barge) systems, the load carrying vehicles (railroad cars or barges) can only move if there are adequate propulsion units (locomotives or towboats). In such cases additional relationships exist to ensure that the total vehicle-hours of those types of vehicles is sufficient to propel the load-carrying vehicle movements. The ratio of propulsion vehicles to load-carrying vehicles is set on each link of the system for each vehicle type to reflect their particular characteristics.

Consumable Resource Limitations (Type G)

Consumables refer to resources that are used essentially continuously in the operation of the system, and include fuel, maintenance supplies, labor, and such items. These must be available in sufficient quantities to support the movement; otherwise vehicle and hence cargo flows will be constrained. Taking the example of fuel, this is accomplished through the use of a fuel consumption estimating relationship for each link of the system and each vehicle type, and using these to estimate total fuel consumption as a function of total vehicle flow. Relationships of Types C and D then make the connection back to the traffic flow. Similar relationships can be developed for maintenance supplies, for labor of various types, as well as other consumables.

Level of Service Constraints (Type H)

The final type of relationship deals with level of service and practical capacity considerations. Most typically these would be used to impose lower limits on the level of service (e.g., maximum O-D transit times). As such, these are optional, and would be employed only when a practical rather than ultimate capacity is desired. Basically these relationships estimate one or more measures of cost as a function of flows (of traffic, vehicles, etc.) and require that costs not exceed upper bounds.

Example of the Model

An example of one version of the model (labeled MAXCAP for later reference) is the following. It is for the situation in which it is desired to determine the capacity of a system, and includes a level of service requirement. The model would include the following relationships, which are presented in the Kuhn-Tucker format of a mathematical programming model:

MAXCAP

Maximize System Traffic (Type A)

Subject to:

Traffic Patterns (Type B)—requirements

Arc Vehicle and Cargo Flows (Type C)

Link Congestion and Delay Relationships (Type D)—limiting vehicle flows

Terminal Capacity and Delay Relationships (Type E)—limiting vehicle and cargo flows, and vehicle storage

Fleet Size (Type F)—limiting traffic

Consumable Resource Limitations (Type G)—limitations given available labor pool, fuel, etc.

Level of Service Constraints (Type H)—minimum LOS

The complete set of model variables, parameters, and equations is presented in Appendix A for freight transportation systems. Also discussed there are variations in the model form to reflect different application contexts and problem types. Appendix B presents the model for passenger applications. While both models are similar, there are many differences in variable and parameter definitions and some differences in relationships.

APPLICATION

The model presented above was applied to an actual network in order to provide a test of the concepts and formulation, and also to provide specific model results that then could be assessed with respect to the usefulness of the results of a model of this type.

Selection of an application was difficult. The reason is that this type of system model requires data on many distinct aspects of the system, including origin-destination traffic volumes, capacities of terminals and links, fleet size and utilization measures, among others. Freight transportation companies generally do not make such data available, even though most of it likely resides somewhere in the organization. Also, some data are sensitive from a competitive standpoint, and others like the facility capacity is not normally called for. In addition, it was desired to apply the model to a multi-firm context, further exacerbating the difficulty.

Test System and Inputs

Fortunately, most of the requisite data was found for a portion of the railroad network. This is the double-stack network service that was the subject of a major study of market potential covering the entire U.S. and most major railroads (Smith et al., 1990). This study, by Manalytics, estimated potential origin-destination flows, thus providing the traffic volumes necessary for the traffic pattern vector and the comparison of capacity with a base case of estimated 1987 traffic (representing "current" volumes). Capacities of many of the major terminals were provided in the study, as were possible train volumes for double stack service on many major routes. The double stack operation is relatively distinct from that which provides other rail services so this created a relatively realistic application. Trains carrying double stack container traffic do not carry regular carload freight, and often contain only double stack cars. Also, the terminals used are exclusively for intermodal containers or trailers, and there is essentially no use by double stack trains of regular rail classification yards. Rail links are, of course, jointly used.

The specific application was to a portion of the 1987 double stack network, the portion shown in Figure 3. Its scope was limited partly by unavailability of relevant data for other parts of the network, and also by a desire to keep this application small in order to keep the results manageable. Major inputs to the model are presented in Tables 1 and 2, which present the traffic pattern matrix, the capacities of the terminals, and the (uncongested) link transit times. It should be noted that the times in Table 2 are not cycle times. Cycle times include terminal times, and in the case of containers, allowance must be made for time consumed while being transported to and from shippers and consignees, time spent loading and unloading, etc. The fleet size is 2,883 cars. It is assumed that the locomotive fleet is adequate, so no constraint for locomotives is included. Traffic flows beyond this network, such as flows to New York, are excluded. No restrictions on level of service (Type F relationships) or on consumables (Type E) were included in this application. (Details of the model are presented in Appendix A.)

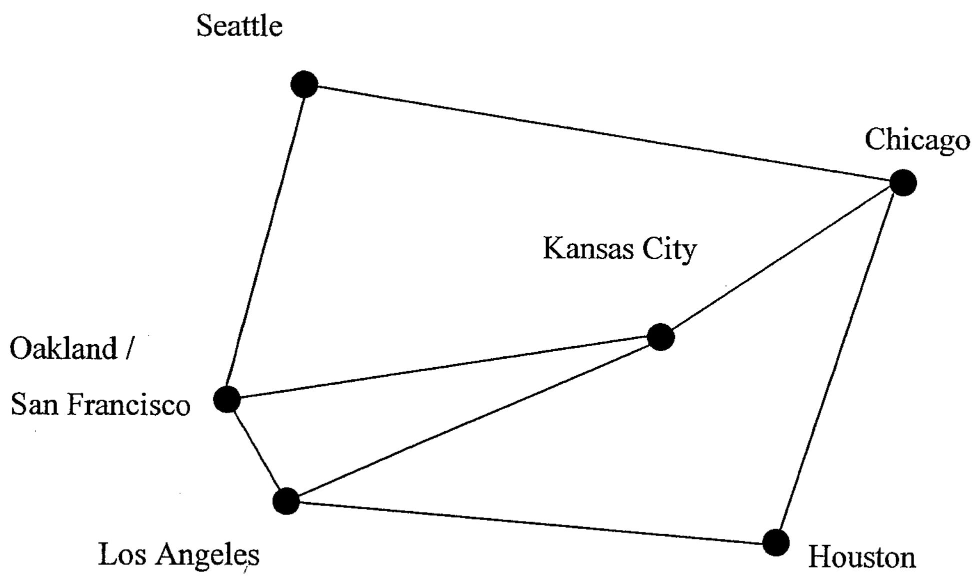

FIGURE 3. The Test Network

Results

Various outputs corresponding to answers to different questions regarding capacity can be obtained from application of this model. Examples of two will be presented in this paper. One is the capacity of the system. The second is a determination of where investments need to be made in order to increase its capacity.

System Capacity with Current Trade Pattern

The MAXCAP version of the model was applied to estimate the capacity of the current system. It is estimated to be 258,717 containers (FEUs) originated per month, in contrast to a current (base case) volume of 201,155 containers per month. This represents a 28.6% increase. In the context of growth in domestic rail traffic of this class of about 6% in recent years, and growth of imports and exports of 2.6% recently, this represents a healthy but not enormous slack capacity.

Identifying Needed Capacity Expansion

In addition to estimating the capacity of the existing system, the model can also determine where increases in capacity must be made in order to accommodate increased traffic. This expansion of capacity can be through investments in assets such as fleet and facilities, or increases in resource allocations such as fuel, if applicable (as they were during the Oil Embargo). This is done using another version of the model, termed ADDCAP (for ADDitional Capacity). This model is presented below, with notations for the specific form used in an analysis which focused on the question of what investments are needed to increase capacity.

The model is:

ADDCAP

Minimize: Facility Capacity (Type C) and Fleet (Type D) Additions—additions required

Subject to:

System Traffic (Type A)—specified level (of increase)

Traffic Patterns (Type B)—requirements

Arc Vehicle and Cargo Relationships (Type C)

Link Capacity and Delay Relationships (Type D)

Terminal Capacity (Type E)—incorporating possible capacity additions

Fleet Size (Type F)—incorporating possible fleet additions

Consumable Resources (Type E)—not included

Level of Service Constraints (Type F)—not included

A very useful way of visualizing the results of such analysis is shown in Figure 4. This figure presents increases in traffic from the base level on the horizontal axis. On the vertical axis is the percentage increase in capacity or number of resources needed. The lines indicate which resources or assets need expansion, and at what traffic level. Specifically, the existing system needs no expansion up to 258,717 FEUs (containers) per month. At that point, the binding or limiting element of the system is the car fleet, indicated by the sloped line starting at that point on the x-axis. This line gives the magnitude of increase in the fleet needed to achieve higher capacities. For example, in order to accommodate 300,000 FEUs per month, then the fleet would have to be expanded by 16.0% (the precise figure is available from the model output), and if 400,000 were the target, then the car fleet would have to grow by 54.6%. To accommodate this latter traffic figure, the capacity of the Seattle terminal will also have to be expanded, by 19.1%. Other facilities and resources are adequate up to just above 400,000 FEUs per month, but at various points beyond that other facilities must be expanded to accommodate additional traffic, as revealed by the other lines in Figure 4. (Note that the lines for capacity expansion at Los Angeles and Oakland/San Francisco are essentially coincident and thus appear as a single line on the figure.)

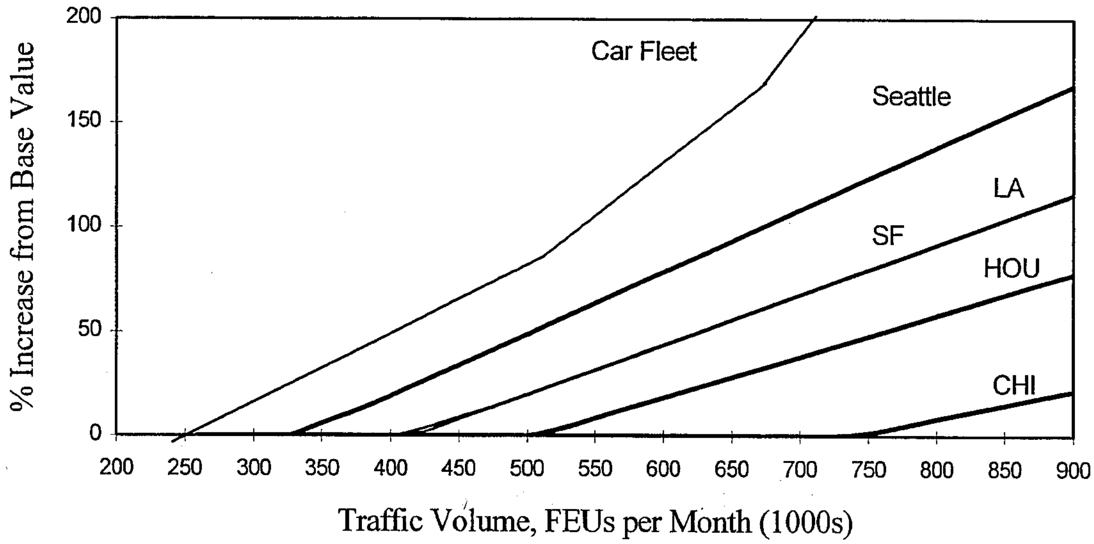

FIGURE 4. Additional Fleet and Terminal Capacity Needed (% of Base Level) to Accommodate Various Levels of Traffic (FEUs per Month).

The non-linearity in the upper end of the fleet curve in Figure 2 reflects the effect of link and terminal congestion. This results in a more than proportional increase in car-hours required as traffic increases, due to increases in the running time of trains and terminal delays. Other sources of non-linearity are also present in the model, such as changes in the pattern of empty car repositioning as traffic increases resulting from added delays or reaching link or terminal capacity limits on least time/cost routes.

Other Analyses and Outputs

Other analyses and outputs can be undertaken with this model, although space limitations permit only a brief review of them here. A particularly important option is to consider changes in the trade pattern and its effect on capacity. For instance, in the example above the predominant flows are east-west, while with the signing of the NAFTA agreement north-south traffic is expected to grow more rapidly than most other international trade. (In the year between April 1994 and March 1995, the dollar value of trade between the U.S. and Canada grew by 15%, and trade with Mexico also increased by 15% (U.S. Department of Transportation, 1996).) The effect of this can be explored in a variety of ways. One would be to use a new trade pattern that reflects the expected future pattern. Another is to incorporate relationships in the model that increase selected flows more rapidly than others, e.g., north-south would increase at 1.7 times the rate of increase in east-west flows, using directly forecast growth rates of each particular O-D traffic flow that could be obtained from economic activity and trade forecasts. The results can be presented in a manner similar to Figure 4, for easy visualization.

An identical type of analysis is applicable to urban areas, where there are pressing problems with congestion affecting both freight and person movements. Much of the problem is that cargo and person movements is increasing rapidly, due in large measure to the suburbanization of residences and commerce, which results in rapid increases in truck-miles and automobile-miles. The ability of the system to accommodate continued increases with or without planned expansion of roads and transit lines can be assessed. This then can provide a basis for either changes in transport plans or changes in development patterns so that they are consistent with transport capabilities. And in many cities, from New York to Beijing, transit systems are taxed to their limit, and this model would provide useful information on the upper limits of passenger traffic growth before major investment is necessary.

Another type of analysis would be to examine the effect of technological advances on the capacity of the system. The ITS program and others, such as Advanced Train Control programs in the railroad industry, will significantly affect the capacity of links and terminals. An important question is what effect these will have on the capacity of the entire system. This can be explored by incorporating increases in facility capacity and shifts in the link delay functions on those lines that are likely sites of such investment, and running the model to determine the effect on system capacity. The inter-relationship between elements that are necessary to achieve increases in capacity can thus be determined (e.g., the effect of the combination of increased rail traffic and lower turnaround time on fleet requirements), just as they were in the previous numerical example.

The model can also be used to assess the effect of catastrophes and other unusual circumstances on system capacity. One example is an earthquake or flood that severs lines of many transport companies, as occurred last year. The ability of the system to support evacuation of citizens is critical under such circumstances. Also, after some transportation links have been severed, the ability of the rest of the system to handle necessary traffic is important. Here a version of the model that explores and identifies good ways of operating remaining links is important, as in the case of the new commuter rail service in Los Angeles responding to the loss of freeway capacity by quickly expanding routes and train services.

Another important application is to military preparedness issues. The same tightening of capacity to handle civilian cargo on the freight system will also limit military movements in time of crisis. Thus the model can be used to assess the system's ability to accommodate a sudden increase in military traffic, as occurred during Desert Storm. Specific traffic additions could be obtained from the military conflict scenarios used in preparedness planning. Also the effect of fuel or other shortages (consumables) can be determined.

Other application contexts and problems could be mentioned, but this list serves to indicate the generality of this type of model to deal with various problem scenarios.

CONCLUSIONS

The motivation of this work was to develop a way of quantitatively assessing the capacity of a transportation system. Prior research was examined, and sound ideas which could be augmented and then translated into a quantitative model for transportation applications were found. A mathematical programming model was developed and then a test of the model was performed using data from a portion of the U.S. freight rail system. The model appears to yield reasonable results. Thus a sound initial model for assessing system capacity appears to have been developed. This model has been developed to the point where it could be applied to a variety of situations, possibly with minor modifications for the specific mode or modes in question in order to capture their unique characteristics. The next step would be to undertake such other applications, either in the U.S. or abroad, where capacity issues are also very pressing, particularly in developing nations. This is contemplated.

REFERENCES

Cambridge Systematics, Inc. (1998) *Multimodal Corridor and Capacity Analysis Manual.* National Cooperative Highway Research Program Report 399, Transportation Research Board, National Research Council, Washington, D.C.

DeLeeuw, F. (1962) The Concept of Capacity. *The American Statistical Association Journal* (December 1962), pp. 826-839.

Eno Transportation Foundation, Inc. (1996). *Transportation in America—Historical Compendium 1939-1995.* Lansdowne, VA.

Eno Transportation Foundation, Inc. (1998) *Transportation in America 1997.* Lansdowne, VA, p. 64.

Kaplan, P. (1997) Grappling with Gridlock; Shippers Jeer Freight-Rail Marriage. *The Washington Times*, December 3, p. B7.

Khan, A. M. (1979) *Railway Capacity Analysis and Related Methodology.* Report No. 60-79-01, Canadian Transport Commission. Research Branch. Ottawa/Hull.

Kneafsey, J. T. (1975) *Transportation Economic Analysis.* Lexington Books, Lexington, MA.

Morlok, E. K. (1978) *Introduction to Transportation Engineering and Planning.* McGraw-Hill, New York.

Morlok, E. K. (1981) Determining The Capacity Of A Transportation System. *1981 Annual Report of the Institute of Applied Systems Analysis*, Ministry of Research and Technology, Federal Republic of Germany, Cologne, pp. 253-260.

Nijkamp, P., A. M. Vleugel, and E. Kreutzberger (1993) Assessment of Capacity in Infrastructure Networks: A Multidimensional View. *Transportation Planning and Technology*, Vol. 17, pp. 301-310.

Scheppach, R. C., and J. G. Faucett. (1973) *Transportation Capacity.* Report No. JACKFAU-73-72(2). U. S. Department of Transportation.

Schipper, Lee et al. (1997) Energy Use and Carbon Emissions from Freight in 10 Industrialized Countries: An Analysis of Trends from 1973 to 1992, *Transportation Research-D*, Vol. 2, No. 1, pp. 57-75.

Smith, D. S. et al., Manalytics, Inc. (1990) *Double Stack Container Systems: Implications for U.S. Railroads and Ports.* Report DTFR53-88-C-00020. FRA and MARAD, U.S. Dept. of Transportation.

Transportation Research Board (1994) *Highway Capacity Manual, 3rd Ed.* (1994) SR209, National Research Council, Washington, D.C.

U. S. Department of Transportation (1996) *Transportation Statistics Annual Report 1996*, Bureau of Transportation Statistics, pp.19-24.

U. S. Department of Transportation (1997) *Transportation Statistics Annual Report 1997*, Bureau of Transportation Statistics, pp. 6, 14, 193-222.

APPENDIX A

Freight Transport Model

MAXCAP - Freight

Model Relationships

All elements of the model are described in detail below. The presentation is organized around the equations, which correspond to the word description presented previously, which is repeated below for convenience in referring to it. The equations are presented together in Table A-1, and all parameters and variables are defined in Table A-2. Units are given, but of course these can be changed easily to suit different applications (e.g., from tons/yr. to tons/day).

MAXCAP

Maximize System Traffic (Type A)

Subject to:

Traffic Patterns (Type B)—requirements

Arc Vehicle and Cargo Flows (Type C)

Link Congestion and Delay Relationships (Type D)—limiting vehicle flows

Terminal Capacity and Delay Relationships (Type E)—limiting vehicle and cargo flows, and vehicle storage

Fleet Size (Type F)—limiting traffic

Consumable Resource Limitations (Type G)—limitations given available labor pool, fuel, etc.

Level of Service Constraints (Type H)—minimum LOS

A. System Traffic (Eqtn. 1)

System traffic (TOTCARGO in the model) is measured by total traffic in tons/yr, or a similar unit. When determining system capacity, this is the quantity that is maximized in eqtn. 1 .

B. Traffic Pattern (Eqtn. 2)

The traffic flows from each origin to each destination (LANE) designated CARGO (LANE), are always subject to the traffic pattern constraints which ensure that an appropriate amount of the

TABLE A-1. Maximum Capacity (MAXCAP) Model Formulation for Freight Transport

A. Maximize Total System Cargo Traffic

$$\text{Maximize TOTCARGO} = \Sigma \text{ CARGO (LANE)} \quad (1)$$

Subject to:

B. Traffic Pattern:

$$\text{CARGO (LANE)} = \text{ALPHA (LANE)} \bullet \text{TOTCARGO} \quad (2)$$

C. Arc Vehicle and Cargo Flows:

$$\text{AVLD (LANE)} \bullet \text{LDCNTR (LANE)} = \text{CARGO (LANE)} \quad (3)$$

$$\text{CARVOL (ARC)} \geq (\text{LDCNTR (LANE)} + \text{MTYCNTR (LANE)}) / \text{CARCAPY} \quad (4)$$

$$\text{TRAINVOL (ARC)} \geq \text{CARVOL (ARC)} \bullet \text{CARLNTH} / \text{MAXTRAINLNTH (ARC)} \quad (5)$$

D. Link Congestion and Delay

$$\text{CARTIME (ARC)} = \text{DELAY [TRAINVOL (ARC)]} \quad (6)$$

$$\text{CNTRTIME (LANE)} = \Sigma \text{ CARTIME (ARC)} + \text{TERMTIME(LANE)} \quad (7)$$

E. Terminal Capacity

$$\Sigma \text{ (LDCNTR (LANE)} + \text{MTYCNTR (LANE))} \leq \text{LIFTCAP (TERM)} \quad (8)$$

F. Fleet Size

$$\Sigma \text{ (}\Sigma \text{ CARTIME (ARC)} \bullet \text{CARVOL (ARC))} \leq \text{CARFLTHRS} \quad (9)$$

$$\Sigma \text{ CNTRTIME (LANE)} \bullet \text{(LDCNTR (LANE)} + \text{MTYCNTR (LANE))} \leq \text{CNTRFLTHRS} \quad (10)$$

$$\Sigma \text{ CARVOL (ARC)} = 0 \quad (11)$$

$$\Sigma \text{ (LDCNTR (LANE)} + \text{MTYCNTR (LANE))} = 0 \quad (12)$$

G. Consumable Resource Limitations

$$\Sigma \text{ FUEL (ARC)} \bullet \text{CARVOL (ARC)} \leq \text{MAXFUEL} \quad (13)$$

H. Level of Service

$$\text{CNTRTIME (LANE)} \leq \text{MAXTIME (LANE)} \quad (14)$$

TABLE A-2. Definitions of Variables and Parameters for MAXCAP - Freight Model

ALPHA(LANE) = traffic pattern fraction

AVLD(LANE) = average loaded container load, tons/container

CARCAPY = container capacity of a car, containers/car

CARFLTHRS = car fleet hours available, car-hrs/yr

CARGO(LANE) = cargo volume in lane LANE, tons/yr

CARLNTH = average car length, ft/car

CARTIME(ARC) = line haul time for car on ARC, hrs.

CARVOL(ARC) = car volume on arc ARC, cars/yr

CNTFLTHRS = container fleet hours available, container-hrs/yr

CNTR TIME(LANE) = container cycle time on LANE, hrs

DELAY[TRAINVOL(ARC)] = time delay function for trains on ARC

FUEL(ARC) = average per car fuel consumption on arc ARC, gallons/train

LD CNTR(LANE) = loaded container volume on LANE, containers/yr

LIFTCAP(TERM) = lift capacity at terminal TERM, containers/yr

MAX FUEL = fuel budget, gallons/yr

MAX TRAIN LNTH(ARC) = maximum train length on ARC, ft

MTYCONTR(LANE) = empty container volume on LANE, containers/yr

TERM(LANE) = cycle time on LANE minus $\sum$ CARTIME(ARC) -- the line haul time on LANE, hrs.

TOT CARGO = total cargo transported, tons/yr

TRAINVOL(ARC) = train volume on ARC, trains/yr total traffic TOTCARGO is flowing in each of the traffic lanes. The traffic pattern (described by the ALPHA (LANE)s) can vary depending upon the specific application and problem.

C. Arc Vehicle and Cargo Flows (Eqtns. 3,4, and 5)

These relationships translate lane (O-D) cargo flows into vehicle flows on arcs and, where applicable, container flows. Eqtn. 3 uses average cargo tons per container (AVLD (LANE)) to yield loaded container volume on each lane, LDCNTR (LANE). Then this plus empty container repositioning flow, on each lane, MTYCNTR (LANE) is divided by car capacity (containers/car, CARCAPY) to yield car volume on each arc CARVOL (ARC) (eqtn. 4). Then eqtn. 5 translates CARVOL into train volume on each arc, TRAINVOL (ARC), using average car length (CARLNTH) and train length limits (MAXTRAINLNTH (ARC)). Of course, in some situations it might be desirable to use a shorter train length than the absolute maximum, and then such a value would be substituted.

D. Link Congestion and Delay Relationships (Eqtns. 6, 7 and 8)

These relationships estimate the congestion delay on line haul arcs and lanes. In this application, this is written in terms of car times (CARTIME (ARC)) and container times (CNTRTIME (LANE)) separately, as containers incur some additional terminal delay. The latter is assumed fixed as this is written, though it could be variable, as discussed under item *E* below.

E. Terminal Capacity and Delay (Eqtn. 10)

Terminal capacity is represented as an upper limit on lifts (LIFTCAP (TERM)) at each terminal, although in some applications this could involve a time delay function as well. The exact relationship will naturally vary with the particular mode and type of facility being represented. Of course, a pure delay function that increases the time that is required for terminal processing could be included also, as could an upper limit on container and car storage capacity.

F. Fleet Size (Eqtns. 9 through 12)

Sufficient vehicles must be available to accommodate the traffic, or conversely the traffic carried must not exceed the limitations of available fleet operations. The basic relationship requires that the total time vehicles are available in the period considered (e.g., week, year), measured by vehicle-hours, must be at least as large as the vehicle-hours of vehicle use, whether loaded or empty, and moving or at rest (as at terminals, intermediate nodes, etc.). Eqtn. 9 sums the product of average car time (loaded and empty, both included in CARVOL (ARC)) and car volume, on each arc, and requires that this be no greater than available car-hours/yr, CARFLTHRS (i.e., cars available $\times$ hours in year). A similar equation exists for the containers (eqtn. 10).

A conservation of vehicle flow equation is needed at each node in the system, specifying that the total number of vehicles of each type entering the node (loaded or empty), must equal the total number of vehicles leaving that node (loaded or empty). One set of equations of this type is needed for each type of vehicle unit considered in the model, in this case cars (eqtn. 11) and containers (eqtn. 12). If desired, other vehicle types could be included, such as locomotives, tow-boats, truck tractors, etc., as appropriate for the application and level of detail sought.

G. Consumable Resources (Eqtn. 13)

Consumables refer to resources that are used essentially continuously in the operation of the system, and include fuel, maintenance supplies, labor, and such items. These must be available in sufficient quantities to support the movement. Taking the example of fuel, a fuel consumption estimating relationship would be included for each arc of the system and each vehicle type, using CARVOL(ARC) and possibly other flow variables to estimate total fuel consumption as a function of vehicle flow. Here this is done using average fuel consumption per car on each arc, FUEL(ARC), multiplied by CARVOL(ARC). The sum of these is limited by the fuel budget, MAXFUEL. Similar relationships can be developed for maintenance supplies, for labor of various types, as well as other consumables.

H. Level of Service Constraints (Eqtn. 14)

Most typically these would be used to impose bounds on the level of service (LOS) supplied, e.g., maximum O-D or lane transit times, MAXTIME (LANE) in eqtn. 15. These are optional, but would be employed when a practical rather than ultimate capacity is desired, or when meeting a minimum LOS is important (as with time-definite deliveries).

Changes for Application to Non-rail and Non-containerized Modes (Technologies)

Rail container service was used as the example for these equations because it represents the most complex of the modes, in the sense of nesting cargo in different units for movement (i.e., cargo into containers, then onto cars, and finally cars onto trains). The simplification of the model for other modes and the non-containerized context is rather straightforward. For example, if cargo were carried directly in rail cars or other vehicles (and not in containers), as in the case of most bulk commodities and many other products, then the two equations would be combined, and LDCNTR(LANE) could be replaced by LDCAR(LANE). Eqtn. 4 would then be definitional for cars, or trucks, etc.

If there were no vehicles, as in the case of pipelines, then all car, train, and vehicle fleet variables would be eliminated, and CARGO(LANE) would enter directly into the link and terminal capacity equations (6 through 8). The fleet size relationships would also be suppressed. CNTRTIME(LANE) would be replaced by CARGOTIME(LANE) for the delay and LOS relationships (eqtns. 7 and 14). A multi-modal model would simply combine different equation forms to represent the different routing options (modes) in the network.

Model Size

The critical elements of model size for solving linear and non-linear programs is usually the number of constraints. This characteristic of the MAXCAP - Freight model can be estimated using the following relationship:

$$4 \cdot \text{NLANEs} + 3 \cdot \text{NTERMs} + 3 \cdot \text{NARCs} + 3$$

where NLANEs is the number of lanes, etc.

Thus, to provide a few examples:

NLANEs	NTERMs	NARCs	No. of Eqtns.
90	10	30	483
450	50	150	2403
6000	80	250	24,993
9900	100	300	40,803

Clearly the number of LANEs is the crucial determinant of model size. This is expected, as the number of origin-destination pairs increases as the square of the number of traffic generating areas (TERMs).

APPENDIX B

Passenger Transport Model

MAXCAP - Passenger

Model Relationships

As in the case of the model for freight applications, the model is presented in the form that is most complicated in terms of the production process. In the passenger case, this is for railroad trains with a variable number of cars. Simplification for technologies where vehicle size (capacity) is fixed is straightforward, and is discussed at the end. Also discussed there are changes necessary to incorporate the automobile mode (and other private vehicle modes). Again for ease of understanding on the part of the reader, we repeat the overall structure of the model in words, and then present the complete equations. This discussion refers to Tables B-1 and B-2.

MAXCAP

Maximize System Traffic (Type A)

Subject to:

Traffic Patterns (Type B)—requirements

Route Vehicle and Passenger Flows (Type C)

Link Congestion and Delay Relationships (Type D)—limiting vehicle flows

Terminal Capacity and Delay Relationships (Type E)—limiting vehicle and passenger flows, and vehicle storage

Fleet Size (Type F)—limiting traffic

Consumable Resource Limitations (Type G)—limitations given available labor pool, fuel, etc.

Level of Service Constraints (Type H)—minimum LOS

A. System Traffic (Eqtn. 1)

System traffic (TOTPAX in the model) is measured by total traffic in passengers/yr, or a similar unit. When determining system capacity, this is the quantity that is maximized in eqtn. 1.

TABLE B-1. Maximum Capacity (MAXCAP) Model Formulation for Passenger Transport

A. Maximize Total System Passenger Traffic

$$\text{Maximize TOTPAX} = \Sigma \text{ PAX(LANE)} \tag{1}$$

Subject to:

B. Traffic Pattern:

$$\text{PAX(LANE)} = \text{ALPHA (LANE)} \bullet \text{TOTPAX} \tag{2}$$

C. Route Vehicle and Passenger Flows:

$$\Sigma \text{ PAX(ROUTE, LANE)} = \text{PAX(LANE)} \tag{3}$$

$$\text{CAROCC (ROUTE)} \bullet \text{LDCAR (ROUTE)} \geq \Sigma \text{ PAX(LANE, ROUTE)} \tag{4}$$

$$\text{CARVOL (ROUTE)} \geq \text{LDCAR (ROUTE)} + \text{NONREVCAR(ROUTE)} \tag{5}$$

$$\text{TRAINVOL (ROUTE)} \geq \text{CARVOL (ROUTE)} \bullet \text{CARLNTH} / \text{MAXTRAINLNTH (ROUTE)} \tag{6}$$

D. Link Congestion and Delay

$$\text{TRAINTIME (ARC)} = \text{ARCDELAY } [\Sigma \text{ TRAINVOL (ROUTE)}] \tag{7}$$

$$\text{CARTIME (ROUTE)} = \Sigma \text{ TRAINTIME (ARC)} + \Sigma \text{ TERMTIME (TERM)} \tag{8}$$

E. Terminal Capacity

$$\Sigma \text{ TRAINVOL (ROUTE)} \leq \text{TRAINCAP (TERM)} \tag{9}$$

$$\text{TERMTIME(TERM)} = \text{TERMDELAY}[\Sigma \text{ TRAINVOL(ROUTE)}] \tag{10}$$

$$\Sigma \text{ (CARVOL (ROUTE)} + \text{NONREVCAR(ROUTE)) AVSTOR(ROUTE, TERM)} \leq \text{STORCAP (TERM)} \tag{11}$$

$$\Sigma \text{ PAX(ROUTE, LANE)} \leq \text{PAXCAP(TERM)} \tag{12}$$

F. Fleet Size

$$\Sigma \text{ (}\Sigma \text{ CARTIME (ROUTE)} \bullet \text{CARVOL (ROUTE))} \leq \text{CARFLTHRS} \tag{13}$$

$$\Sigma \text{ CARVOL (ROUTE)} = 0 \tag{14}$$

G. Resource Limitations

$$\Sigma \text{ FUEL (ROUTE)} \bullet \text{CARVOL (ROUTE))} \leq \text{MAXFUEL} \tag{15}$$

H. Level of Service

$$\text{TRAINTIME (ROUTE)} \leq \text{MAXTIME (ROUTE)} \tag{16}$$

TABLE B-2. Definitions of Variables and Parameters for MAXCAP - Passenger Model

ALPHA(LANE) = traffic pattern fraction

ARCDELAY[Σ TRAINVOL(ROUTE)] = time delay function for trains on ROUTE

AVSTOR(ROUTE, TERM) = average total storage and servicing time of a car at TERM, by ROUTE, hours/car

CARCAPY = passenger capacity of a car, passengers/car

CARFLTHRS = car fleet hours available, car-hrs/yr

CARLNTH = car length, ft/car

CAROCC(ROUTE) = average car occupancy of loaded car on ROUTE, passengers/car

CARTIME(ROUTE) = total train run time for ROUTE, hrs

CARVOL(ROUTE) = car volume on arc ROUTE, cars/yr

FUEL(ROUTE) = average per car fuel consumption on arc ROUTE, gallons/train

LD CAR(ROUTE) = revenue car volume on ROUTE, cars/yr

MAX FUEL = fuel budget, gallons/yr

MAX TRAIN LNTH(ROUTE) = maximum train length on ROUTE, ft

NONREVCAR(ROUTE) = non-revenue car volume on ROUTE, cars/yr

PAX(LANE) = passenger volume in lane LANE, tons/yr

PAX(ROUTE, LANE) = passenger LANE volume using route ROUTE, passengers/yr

PAXCAP(TERM) = passenger throughput (boarding and alighting) capacity at terminal TERM, passengers/yr

TERMDELAY(TERM) = time delay function for dwell (loading and alighting) at terminal

TERMTIME(ROUTE) = terminal time for loading and unloading, hrs

TOTPAX = total passengers transported, passengers/yr

TRAINCAP(TERM) = maximum train throughput of terminal TERM, trains/yr

TRAINTIME(ARC) = travel time for car on ARC, hrs.

TRAINVOL(ROUTE) = train volume on ROUTE, trains/yr

B. Traffic Pattern (Eqtn. 2)

The traffic flows from each origin to each destination (LANE) designated PAX(LANE), are always subject to the traffic pattern constraints which ensure that an appropriate amount of the total traffic TOTPAX is flowing in each of the traffic lanes. It should be noted that passenger traffic could be further subdivided into categories, e.g., first class vs. coach. Corresponding subdivision of vehicle capacity would then be necessary (i.e., seating accommodation or car type).

C. Route Vehicle and Passengers Flows (Eqtns. 3 through 6)

Many O-D pairs and hence LANES can use the same ROUTE. This is well illustrated by a train on the New York to Washington ROUTE, which makes intermediate stops at Philadelphia, Wilmington, and Baltimore. Passengers from New York to Washington will occupy the train at the same time that travelers from New York to Baltimore and from Philadelphia to Washington occupy its capacity. Therefore it is necessary to divide the passenger flow into flow by ROUTE, and then apply the relationships which yield car and train volume by ROUTE (rather than ARC as in the case of the freight model). Eqtn. 3 performs this division into PAX(ROUTE,LANE).

Eqtns. 4 through 6 translate passenger flows into car and train flows on routes. Eqtn. 4 uses average car occupancy in passengers per car (CAROCC(ROUTE)) to yield passenger-carrying (loaded) car volume on each route, LDCAR(ROUTE). This equation must hold for each arc on a route, and would be written for all areas where capacity might be binding. Different practices among modes lead to minor variations in exact interpretation. For example, in some contexts revenue (i.e., passenger carrying) car or vehicle flows and occupancy are separate from non-revenue (repositioning or deadhead) movements. In others, data on occupancy ignore such distinctions. Clearly load factors (average or maximum, depending on the application) would be the basis for CAROCC(ROUTE).

These revenue car flows, plus non-revenue car flows NONREVCAR(LANE), on each route, are summed over the appropriate set of ROUTES to yield car volume on each arc CARVOL (ROUTE) (eqtn. 5). Then eqtn. 6 translates CARVOL(ROUTE) into train volume on each route, TRAINVOL(ROUTE), using average car length (CARLNTH) and train length limits (MAXTRAINLNTH (ROUTE)).

D. Link Congestion and Delay Relationships (Eqtns. 7 through 9)

These relationships estimate the total time including congestion delay on arcs, written in terms of train times (TRAINTIME (ARC)).

E. Terminal Capacity and Delay (Eqtns. 9 through 12)

Terminal capacity is represented by four equations. One is an upper limit on train volume, TRAINCAP(TERM). A second (10) yields the terminal processing times for passenger carrying trains, reflecting congestion delays resulting from the train volume (which presumably also reflects and incorporates the passenger volume—or can be expanded to include this). The third limits car dwell time, using average storage and servicing times AVSTOR(TERM) for routes terminating at the terminal. The fourth limits passenger flow. The exact relationship will naturally vary with the particular mode and type of facility being represented.

F. Fleet Size (Eqtns. 13 and 14)

Sufficient vehicles must be available to accommodate the traffic, or conversely the traffic carried must not exceed the limitations of the available fleet. The basic relationship requires that the total time vehicles are available in the period considered (e.g., week, year), measured by vehicle-hours, must be at least as large as the vehicle-hours of vehicle use, whether revenue (loaded) or non-revenue (empty), and moving or at rest (as at terminals, intermediate nodes, etc.). Eqtn. 13 sums the product of average car time (loaded and empty, both included in CARVOL (ROUTE)) and car volume, on each route, and requires that this be no greater than available car-hours/yr, CARFLTHRS (i.e., cars available $\times$ hours in year).

A conservation of vehicle flow equation (14) is needed at each terminal in the system, specifying that the total number of vehicles of each type entering the terminal (loaded or empty), must equal the total number of vehicles leaving that node (loaded or empty). One set of equations of this type is needed for each type of vehicle unit considered in the model.

G. Consumable Resources Limitations

Consumables refer to resources that are used essentially continuously in the operation of the system, and include fuel, maintenance supplies, labor, and such items. These must be available in sufficient quantities to support the movement. Taking the example of fuel, a fuel consumption estimating relationship would be included for each arc of the system and each vehicle type, using CARVOL(ROUTE) and possibly other flow variables to estimate total fuel consumption as a function of vehicle flow. This would be constrained by any limitation, e.g., allocation of fuel in times of emergency. Similar relationships can be developed for maintenance supplies, for labor of various types, as well as other consumables.

H. Level of Service Constraints (Eqtn. 16)

Most typically these would be used to impose bounds on the level of service (LOS) supplied, e.g., maximum lane or route transit times, MAXTIME (ROUTE) in eqtn. 16. These are optional, but would be employed when a practical rather than ultimate capacity is desired, or when meeting a minimum LOS is important.

Changes for Fixed Vehicle Sizes

If there were only one size vehicle, then CARVOL(ROUTE) and TRAINVOL(ROUTE) would be identical, and eqtns. 5 and 6 would be combined. Also, if there were different vehicle types with different sizes, then a distinction would be made between different car types in eqtns. 3 through 6, e.g. CAROCC1 vs. CAROCC2, CARVOL1, etc. Corresponding variable substitutions would be necessary in the link congestion, terminal capacity, fleet, resource limitation, and LOS relationships.

Incorporating Automobile Travel

Clearly the model has been presented as if it were applied to common carriers only—air, bus, and rail services, etc. The same equations can be interpreted to permit inclusion of auto travel, or any other private vehicle mode. However, the inclusion of private automobile travel, and in fact all private vehicle travel by any mode, presents a variety of issues that must be addressed before settling for a particular version of the model. These issues arise forcefully in passenger transportation because in developed nations the vast majority of travel is via private automobile (about 85% in North America and Western Europe).

Changes in the Model for the Automobile Option

Turning first to the inclusion of the automobile, the modifications necessary are minimal. Basically it involves defining a subset of ROUTEs as auto routes. These would have some passengers allocated to them, using an equation like eqtn. 3 to specify the fraction or quantity of travelers who could use (or would choose to use) auto. This likely would limit PAX(ROUTE, LANE) for autos directly.

Eqtn. 4 then would translate these auto passenger volumes into automobile volumes, and then these are subject to limitations due to facility capacity through eqtns. 7 and 8 for link capacity and delay. Eqtn. 9 would represent limitations on parking capacity. Most likely eqtn.12 would not apply, given the use of autos by owners only. The resource limitations and level of service limitations could apply, of course.

General Issues for Including Private Vehicle Travel Modes

Now that the means of including the automobile has been covered, it is appropriate to discuss some of the issues related to its inclusion (or not). In some circumstances it might be appropriate to use the model ignoring the private auto or other private modes, at least as travel options. One would be where the analysis is being carried out from the standpoint of a single carrier, such as a transit authority wanting to know its capacity. Similarly, for some purposes it might be useful to know how many persons all the common (public) carriers in a region could accommodate. Of course, in these cases the effect of private vehicles on the delays and capacity available on shared facilities is important, and must be included. Similarly their effect on the availability of fuel or other resources may be important as well.

Another aspect of including the auto and possibly other private modes (boats, aircraft, etc.) is that in some circumstances it may be used like a common or public carrier. Normally the vehicle is used only to transport the owner or owners where they wish to travel, accompanied by others whom they choose. Thus it normally cannot be assigned to transport other persons, or shifted around the network to maximize passenger throughput, as a common carrier vehicle can be. However, in time of severe emergency, it is possible that private vehicles could be assigned to general person movements and used much like common carrier vehicles. If the latter condition applies, then the private vehicle fleet can be considered part of the common carrier fleet, and incorporated in the same manner that such modes are. Thus, in presenting the model, it was assumed that automobiles are used by their owners only.