Capacity Modeling in Transportation:  
A Multimodal Perspective

UCI-ITS-LI-WP-04-3

Minyoung Park ¹  
Amelia C. Regan ²

¹ Institute of Transportation Studies, University of California, Irvine  
Irvine, California 92697-3600, U.S.A., mypark@uci.edu

² Department of Computer Science, Department of Civil and Environmental Engineering  
and Institute of Transportation Studies, University of California, Irvine  
Irvine, California 92697-3600, U.S.A., aregan@uci.edu

July 2004

Institute of Transportation Studies  
University of California, Irvine  
Irvine, CA 92697-3600, U.S.A.  
http://www.its.ucl.edu
Capacity Modeling in Transportation: A Multimodal Perspective

Minyoung Park
Department of Civil & Environmental Engineering
University of California, Irvine
Irvine, CA 92697-3600
Phone: (949) 824-5623
FAX: (949) 824-8385
E-mail: mypark@uci.edu

Anelia Regan
Department of Computer Science and
Department of Civil & Environmental Engineering
University of California, Irvine
Irvine, CA 92697-3600
Phone: (949) 824-1074
FAX: (949) 824-8385
E-mail: aregan@uci.edu

ABSTRACT
This paper provides a conceptual framework for capacity modeling of multimodal transportation networks. We briefly review the evolution of capacity models for use in transportation systems planning and investment, and provide an in-depth discussion of recent advances towards a system-oriented capacity model and their applicability in the multimodal freight transportation context. A logical network capacity model based on bi-level programming is proposed for measuring the capacity of multimodal freight transportation systems.

Key Words: Freight Transportation Planning, Multimodal Transportation Systems, Network Capacity, Bi-level Programming

Word Count: Total 7,549 Words
Submission Date: July 30, 2004
Submitted for Presentation in the 84th Transportation Research Board Annual Meeting and Publication in Transportation Research Record
INTRODUCTION

Freight transportation plays an enormous role in the regional and national economic growth and development. It supports production, trade, and consumption activities by ensuring the efficient movement and timely availability of raw materials, intermediate products and finished goods. Over the years, interest in explicit consideration of goods movement in the transportation planning process has steadily increased. Recently, however, there is a serious concern in urban regions about how to secure adequate capacity to accommodate rapidly growing freight demand. The demand for freight transportation is increasingly competing with passenger transportation, leading to increased congestion. The delays caused by the congestion result in higher costs for all drivers and also supply chain participants -- producers, transporters, retailers and consumers.

Transportation professionals have proposed many solutions to facilitate efficient freight transportation. These range from constructing new facilities to restricting transportation demand. While these approaches can improve capacity, traditional approaches to capacity preservation and expansion of physical facilities have proven to be inadequate, particularly in urban areas, mainly due to the high cost of land use, environmental concerns, physical barriers, and opposition to new infrastructure voiced by local communities. These constraints make new capacity improvements difficult to implement. Furthermore, controlling freight demand is unlikely to be consistent with the goal of continued economic expansion.

In part because of these issues, there is broad recognition of the need for comprehensive multimodal approaches that leverage the competitive advantages of each mode (1, 2, 3). It seems clear that capacity issues should not be examined in relation to separate infrastructure segments but as part of a multidimensional system. For example, inadequate capacity in some modes may be overcome by intermodal substitution and complementarity, i.e., by enhancing the utilization of excess or unused capacity in other modes. The recent emergence of a shuttle rail service for short-haul freight movement, applied in a corridor where the road network is highly congested and existing rail capacity is underutilized serves as an example (4). In this context, public investments in private rail networks could be justified since capacity problems on road network may be alleviated by making adjustments in rail operations without funding new investments in road capacity. In response, recent transportation policies of public agencies are moving in favor of modally balanced transportation systems by encouraging the development of multimodal freight policy, planning, and programming that meet economic development goals in a manner that includes all feasible transportation modes (5).

The changes from modal fragmentation to cross-modal coordination and from system construction to system optimization require public agencies responsible for transportation planning and investment to more precisely identify and measure capacity gaps across all transportation facilities involved in freight transportation processes. A capacity gap may be described simply in terms of potential weakness in the system which can lead to alternative strategies for developing additional capacity. Capacity assessment is essential for identifying and measuring the capacity gaps since the comparison of information or metrics developed from the estimate provides planners with an indication
of which dimensions need attention and the extent of the impact of capacity gaps relative to current and future demand. In addition, from the standpoint of a decision-oriented planning process, incorporation of capacity measures into the planning process provides better accountability for transportation investment decisions, implying that transportation planners can make more cost-effective decisions.

Despite the growing importance of capacity assessment in transportation planning and investment decision-making processes, capacity models for use in transportation development to date appear inadequate for multimodal systems capacity analysis. These are limited either to measuring the capacity of individual elements of systems (for example, the Highway Capacity Manual (6)), or to capturing the capacity of partially aggregated components of systems, i.e., either a corridor (6, 7) or a modal network (8, 9, 10). The latter presents a significant extension beyond the former, but these approaches do not adequately address the distinct features of multimodal systems such as responses of system users to multimodal options, intermodal connectivity, and interactions between many other system elements. From a multimodal systems viewpoint, the capacity problems are a function of a combination of inadequate considerations of broader systems factors, poor integration and coordination of various systems dimensions, and/or dependency on external factors such as land use, economy, environments, and technology. While simultaneous consideration and integration of the capacity requirements and inter-relationships made up by the interactions of multiple actors in the multimodal systems trigger greater complexity in capacity problems, new emphasis needs to be on a much broader systems approach to capacity analysis.

In recognition of this need, this paper proposes such a model, from which a conceptual network model based on bi-level programming is developed to estimate the capacity of multimodal transportation systems. This approach may allow us to articulate investment priorities across all transportation modes and to realize sustainable development objectives through an integrated systematic capacity analysis framework. Building on earlier efforts, new information and advanced methods for capacity analysis are important for better resource allocation and utilization in transportation infrastructure planning and management. Capacity estimates from this framework have direct implications for strategic freight transportation planning, infrastructure investment decision-making, and related policies.

The paper is organized as follows. In the next section, we briefly examine the historical evolution of capacity models. A discussion is provided of recent advances toward integrated, system-oriented capacity models and their applicability in the multimodal freight transportation context. This is followed by the identification of key factors that influence system capacity, from which we define a generalized multimodal network capacity problem in the context of freight transportation. A conceptual framework is then described for formulating network capacity models based on bi-level programming, which provides a basis for measuring the capacity of multimodal transportation systems. Finally, the paper concludes by presenting future research directions.
EVOLUTION OF CAPACITY MODELS IN TRANSPORTATION

The earliest formal effort to evaluate transportation capacity is the Highway Capacity Manual (HCM) published in 1959 (6). Its subsequent editions were published in 1965, 1985, 1994, and 2000, respectively, to reflect advances in analyzing the operational performance of road systems. The HCM presents methods for analysis of capacity and level of service for a broad range of transportation facilities, including urban streets, freeways, (un)signalized intersections, weaving sections and ramps, among many others. Capacity in the HCM is defined as the maximum number of either vehicles or persons that can traverse a point or uniform segment on the highway during a given time period under prevailing roadway, traffic, and control conditions.

In contrast, the first transit capacity manual in the U.S. was published in 1999 (17). While the capacity concepts in the HCM have been widely accepted and applied, previous efforts for capacity analysis have been too focused on transportation via road. This problem was recognized by the latest version of the HCM in part by adding sections dealing with transit capacity, pedestrians, and bicycle facilities (6). Capacity analysis procedures for other modes have been also developed in parallel, but these have evolved around the use of the capacity concept defined in the HCM (see Cambridge Systematics, Inc. et al. (7) for more details). The HCM-type capacity notions primarily measure the capacity of individual facilities of transportation systems such as road segments, railway lines, and terminals. The facility capacity may be suitable for project-level analysis of individual components, but as pointed out by Morlok and Riddle (8), it does not account for the overall capability of the system.

As the policy environment for transportation planning changed during the 1990s towards a modally balanced transportation system, transportation professionals have begun to pay more attention to the development of analytical tools and data to implement multimodal transportation planning process that considers not only all possible modes but also low-cost capacity improvements. Along this line, two new approaches to capacity assessment have been proposed. One is a method for multimodal corridor capacity analysis (6, 7). This method estimates the capacity of a corridor by determining the capacity of the lowest capacity element of each path or of the sum of the capacities on parallel paths. This kind of analysis represents a significant advance in the direction of multimodal systems capacity analysis, but it is limited to the capacity of a partial set of component parts of the entire system.

Another approach associated with capacity estimation is a system-wide capacity analysis. The first work related to the system capacity found in the literature is Morlok and Riddle (8), in which the concept of system capacity is well defined and a quantitative model based on an optimization technique is presented to determine the capacity of a transportation system. The optimization-based model estimates the maximum freight volume that a rail network can accommodate while satisfying a set of constraints including fixed origin-destination traffic pattern and resources. A variant of the model is also developed by permitting the variations in the base traffic pattern to determine where capacity must be added to accommodate increased traffic. These models are used later by the authors to measure the capacity flexibility of transportation systems that can.
accommodate unexpected changes in demand and traffic patterns (12). The system capacity concept was also applied by Yang et al. (9) and Ge et al. (10) to modeling road network reserve capacity. Given a capacitated network configuration, the former considers the zonal trip generation potential and/or level of service constraints under network equilibrium conditions, while the latter addresses the impacts of travel information on network capacity.

The system approach to capacity assessment is noteworthy in terms of extending existing capacity concepts to a system context. Yet even this definition will suffer from some shortcomings when evaluating the capacity of a multimodal freight system. A main drawback of this approach lies in the fact that they do not consider the likely impacts of modal substitution on system capacity. Freight movements depend not only on a modal network capacity, but also on the ability of system users to substitute transportation modes and transfer across networks through intermodal connections. Depending on the availability of multimodal options by network users, we may have different measures of the maximum network flow. Furthermore, they fail to notice the existence of multiple products (or users) that have different service requirements for transportation. Any changes in operations to meet these diverse needs directly have significant impacts on the system performance. Nonetheless, the system capacity concept provides useful insight for enhancing the utilization of existing capacity by evaluating the capacity problems from a system perspective, and becomes a basis towards developing a broader systems capacity concept.

In fact, the failure of many projects and programs that deal with capacity improvement can often be attributed to the narrow view of capacity that had been used. As mentioned in Nijkamp et al. (13), capacity is essentially a multi-faceted phenomenon, meaning that it cannot easily be characterized by means of single indicator, but needs to be viewed from multiple dimensions. This is especially true of the multimodal freight transportation systems, which involve the movement of goods by the combined use of multiple modes and facilities. More functions are added, creating greater complexity in capacity analysis. Therefore previous state-of-art capacity analyses should be extended to characterize the distinct principles of multimodal freight systems. A significant challenge to devising a tool for multimodal systems capacity analysis is the compound and diverse array of entities, issues and needs associated with the capacity problems. Accordingly an effort for assembling and organizing these and prioritizing across them must be a first step towards building a common definition of multimodal network capacity problems and a comprehensive framework for formulating an integrated, systematic capacity model. These issues are addressed in the subsequent sections.

KEY DIMENSIONS OF MULTIMODAL SYSTEMS CAPACITY

Multimodal freight transportation systems consist of numerous components which must perform together for the systems to work successfully. The inherent function of each component and interaction between different components form key dimensions of the systems capacity. In developing an advanced model, it is useful to identify key factors that describe capacity limitation of the system and to understand the relationships between the factors and capacity. While a wide variety of system elements exist, six core
components deserve special attention, named as the VI-CORE model, including: (1) Vehicles, (2) Infrastructure, (3) Customers, (4) Operations, (5) Regulations, and (6) Externalities. Figure 1 illustrates the VI-CORE model that identifies these key components. Each of the basic relationships is briefly described as follow.

![VI-CORE Model Diagram](image)

**FIGURE 1. The VI-CORE Model: Six Core Dimensions of Systems Capacity**

- **Vehicles**: The first major component of multimodal freight systems is vehicles. The vehicles component characterizes variable resources that convey goods from place to place, such as truck, rail, airplane and ship. They have different functioning characteristics and purposes. For example, most short-haul moves involve trucks, while rails serve long-haul freight. Other modes such as air and maritime can also be a part of the multimodal systems, but are normally limited to international trade or express parcel operations. Distinct characteristics of each mode lead to different perceptions of users, thus leading to unbanned level of usage for each available mode. The disparity in utilization of each mode, in turn, entails different values of systems capacity, depending on the degree of coordination of unexploited vehicles' capacities. The primary determinants of systems capacity in this category are the number of available modes, fleet size of each mode, and loading rate of each vehicle type.

- **Infrastructure**: In contrast to the vehicles component, infrastructure denotes the fixed part of multimodal systems. Highways, railroads, airports, harbors and freight terminals are major elements of this dimension. Successive combination of the facilities defines the feasible paths of freight movement in the multimodal network. The physical condition and related operational performance of each facility are important in maintaining acceptable levels of mobility, leading to different measures of systems capacity. Often overall performance of multimodal systems is attributed to the degree of modal connectivity. Thus special attention should be paid into this component.

- **Customers**: 'Customers' as a systems component is a generic term that characterizes key participants involved in freight transportation processes. In the context of freight
transportation, there are two major user groups who have different roles and objectives. These are shippers and carriers. Shippers represent all the parties involved in the processes of trip-making decisions, destination and mode choices, including producers, wholesalers, retailers, and individual consumers. Freight carriers including trucking companies, railways, shipping and air lines provide transportation services in response to the shippers' demand for transportation. Depending on the type of service they provide, ports, intermodal platforms and other such facilities may be described as carriers as well. They are essentially engaged in the processes of resources allocation and routing and scheduling. One may also include governments and other types of stakeholders - those segments of a community negatively affected by freight transportation such as traffic accidents and air pollution. But considering the importance of these segments, they are treated as independent components in this paper. The aggregated decisions by all shippers and carriers involved create the basic freight flow patterns in multimodal networks that demand for capacity. Thus, the behavioral aspects of individual system users become an important consideration in the systems' capacity assessment.

- **Operations**: The operations component refers to the set of procedures by which individual vehicle- and shipment are guided on the fixed facilities. Unlike passenger transportation, the capacity for goods movement is more difficult to quantify due to its nature dealing with multiple commodities. Each commodity has different attributes in terms of value, weight, size, packaging, and others such as perishable and hazardous products, requiring different levels of transportation services. To be responsive to these diverse needs, vehicles and other resources needed are operated in different ways. Any changes in operations of the resources directly have significant impacts on the systems performance. Therefore precise coordination and optimization in operations of available resources and handling of shipments with similar characteristics for consolidation will improve the utilization of existing capacity in the systems. Service types and frequency, mixture of available resources, vehicles routing and scheduling, consolidation rules, pricing schemes are all examples in this dimension.

- **Regulations**: The regulations component identifies the involvement of public agencies related to the capacity problems. Some of these are connected with their traditional roles, such as new construction of transportation infrastructure, subsidies to private companies to ensure service to mobility-disadvantaged groups, and public ownership of highways to maintain a comprehensive level of accessibility. However, involvement of public agencies does not always entail capacity expansion. By placing legal requirements on individuals and firms to meet the public goals, they also involve those actions that strain the full use of existing capacity. Government regulations arise largely from recognition of public concerns such as safety, environment and equity issues, and have become one of the important forces behind the underutilization of existing capacity. Consequently, investment and regulation actions of public agencies are closely related to the systems capacity by directly and indirectly affecting the operational performance of the whole system, making this segment a key consideration.

- **Externalities**: In fact, the transportation system itself exists within and interacts with other systems such as the economic or ecological system. This implies that capacity issues should be extended further to reflect the impacts of activities in other systems. The externality component demarcates those segments that make up a set of constraints on the
use of infrastructure and resources. From a broader systems perspective, economic development of some regions primarily changes the pattern of spatial freight distribution, resulting in variation in traffic flow pattern in the network. The systems perspective also leads to the observation that increase in freight activities often has negative impacts on society such as traffic accidents and air pollution. Clearly these impacts limit the full use of existing capacity. Thus this dimension must be explicitly addressed.

THE CONTEXT

The systems perspective addressed here has important implication for problem definition, scale of analysis and methods that deal with such broader issues and implementation strategies that transcend traditional boundary. Obviously no one-size-fits-all model can encompass all problems. Our focus is on models aimed at evaluating the multimodal systems capacity and assisting the decision-making process at strategic planning level. One of the first steps in any modeling efforts is to establish the boundary of the systems to be analyzed. It is often necessary to make assumptions suitable for the application contexts, which can be tailored to a wide variety of multimodal practices. Based on knowledge of the basic relationships between systems components and capacity discussed above, we now define the multimodal freight systems capacity problem.

We consider a transportation network which includes the transportation facilities available for freight transportation. They are fixed in space and each facility has the limited capacity. There is demand associated with freight movement which needs to be transported over the network from one place where it is produced to another where consumed. Various forms of transportation services are available to serve the demand, such as truck, rail, and the combination of these modes. They are integrated through intermodal facilities where modal transfer takes place. In this setting, we assume that there is a single decision-making entity which may be a supplier of transportation infrastructure or a controlling agency, whose objective is to make optimal decisions with respect to freight transportation systems planning and investment under its control at a regional or national level. To achieve this goal, the agency wishes to know maximum amount of freight traffic that the integrated network can accommodate, while satisfying some constrained network conditions. The identification of the optimal mix of necessary infrastructure nodes in view of reaching given objective is a major issue in this context.

We refer to this problem as the Multimodal Network Capacity Problem (MNCP). A well-designed, refined capacity model of the MNCP can be used to answer some fundamental questions associated with capacity problems. Some of these are: Where is the weakest point or segments (i.e., bottlenecks) in the network? How can maximum utilization of existing capacity be assured? And, given planned infrastructure investments and other operational changes, how adequate and flexible will the capacity of the system be in the future? In what follows, we describe a conceptual framework for formulating the MNCP model in a way that fully integrates key factors seen to be critical in the context of freight transportation.
THE MULTIMODAL NETWORK CAPACITY MODEL

The mathematical model that best describes the MNCP defined in the previous section is the well-known multi-commodity network flow problem. Problems of this kind are well studied in the literature (see Ahuja et al. [14] for example). The essential issue addressed by the network flow problem is the allocation of the capacity of each link to individual commodities in a way that minimizes overall flow cost on the network. This concept has been used previously in examining transportation problems such as congestion pricing (15), traffic signal control (16) and road network design (17). In the context of multimodal freight transportation, however, the problem becomes quite complex, requiring advanced models that encompass the distinct characteristics of multimodal freight networks. In this section, we present a conceptual model for the analysis of the MNCP to determine optimal capacity measure.

Representation of Multimodal Network

The modeling framework we tackle is essentially that of a multimodal freight network made up of various facilities, on which multiple products are moved by specific vehicles between given origin and destination points. As a convenient means of arranging information about the characteristics of network and freight movements over the network, we introduce an integrated network in which each node involved has its own distinct physical network and these are integrated through intermodal transfer facilities. The following paragraphs present the basic network notation. Further notation is introduced as needed.

The integrated network is represented by a directed graph that includes a set of nodes, \( V \) and a set of directed arcs (links), \( A \). The set of nodes characterizes all product origins and destinations, transfer facilities and the intersections of different line-haul segments. To define an origin-destination (O-D) pair, we distinguish two special nodes in the network: an origin node \( r \in R \subseteq V \) and a destination node \( s \in S \subseteq V \). Each O-D pair \( w = (r, s) \), \( w \in W \) is connected by a set of paths through the network, denoted by \( P_w \). The set of links represents line-haul segments that connect two points in \( N \). Each link \( a \in A \) has its natural attributes such as length, vehicle type allowed, and capacity. Freight volume \( x_a \) and cost \( c_a \) are also associated with each link \( a \). Parallel links are used to model goods movements between two adjacent nodes in \( N \) by different modes. By assigning a separate link to each mode, we can easily identify not only the flow of goods by different modes on the same route, but also different types of services (e.g., express vs. regular) and different carriers in the same mode.

A mode \( \gamma \) mode transfer is an integral part of the integrated network. In order to capture intermodal movements, it is necessary to allow modal transfer at certain nodes of the network and to compute the associated cost and delay. This can be realized by expanding a single node where a modal shift occurs by adding as many nodes as links entering and exiting the node and by adding transfer links between these nodes. The inclusion of the conceptual links may also be useful for defining other terminal activities such as loading, unloading and consolidation in a similar way. In this context, a path
Given the physical network configuration, we now wish to find the maximum flow that the network can accommodate. In multimodal systems, there exist different levels of decision makers, each with different objectives. Basically decision makers at the higher levels solve their problems while taking into account responses of lower levels in the form of constraints. If there is only one top-level decision maker, optimality of the highest level problem can be achieved after lower level problems have reached an optimal (or equilibrium) condition. This is especially true of the MNCP, in which identification of capacity gaps over a multimodal network is the ultimate goal and this can be attained only after assigning current demand to the existing capacity. Problems of this form are known in the literature as bi-level programming problems, which have been used by a number of authors in the transportation area. See for example, Bard (18), for a general overview of this approach. To construct a model of hierarchical systems in general, behavior at the lowest level is modeled first and higher levels are added in steps.

Integrated Network Optimization
Once a network is defined, the problem at hand is to characterize the equilibrium flow pattern on the network. In transportation, network equilibrium is the interface between demand and performance, its prediction is at the core of the analysis of any transportation system. The distribution of freight on the integrated network may be modeled in the framework of conventional network assignment models, in which an equilibrium flow pattern is determined in a way such that an optimality condition is satisfied subject to a set of constraints associated with flow on the network. Cramin (20) present an in-depth review of these studies related to freight transportation.

To describe a possible equilibrium model that may be integrated into the MNCP, let us define the flow and cost elements and their relationships on the network as follow. Freight demand on the network can be specified by an origin-destination matrix, indicating the demand per unit time for transportation services between each O-D pair. This is generally categorized by different classes of products, each of them with distinct demand and cost functions and different valuation for the cost factors. Let \( K \) denote the set of all products considered. A product \( k \in K \) represents any product or collection of similar products that generates flow within the network. The demand of product \( k \) for O-D pair \( w \) is defined by an O-D matrix \( q_{w,k}^{*} \), in which each element may be measured in units of tons per unit time in question. Note that the demand may be fixed or elastic depending on the relationship between demand and costs. At this point in our discussion, it is assumed as fixed and thus exogenous to the model. The latter case will be addressed later. Then we have a set of flow conservation equations, stating that all path flows of the same O-D pair sum to the total O-D flow for each product \( k \), as follows.
\[ \sum_{p \in P} h^k_p = q^k_p, \quad \forall k, p \]  

(1)

where \( h^k_p \) is the flow of product \( k \) on path \( p \). Here \( h^k_p \) must be non-negative to ensure that the solution of the problem is physically meaningful, expressed as

\[ h^k_p \geq 0, \quad \forall p, k \]  

(2)

The flow of all the products on the network is denoted by \( x = (x^1, x^2, \ldots) \), where \( x^a = \sum_k x^k_a \) and \( x^k_a \) is the flow of product \( k \) on link \( a \). By definition, \( x^k_a \) and \( h^k_p \) are related as

\[ x^k_a = \sum_{p \in P} \delta^a_p \cdot h^k_p, \quad \forall a, k \]  

(3)

where \( \delta^a_p \) is an indicator variable that has 1 if link \( a \) is on path \( p \); zero otherwise.

Let \( H \) be the set of feasible path flows that satisfy equations (1)-(3). Given the set of feasible path flows, an optimality criterion must be established to determine a unique, optimal flow pattern. Obviously an optimal flow pattern is a direct consequence of the aggregated decisions made by all individuals in the network. Thus it must be determined after properly specifying the actors involved, their behavioral rules and their relationships. We do not consider shippers and carriers as distinct actors in the decisions made in shipping freight. Instead, it is assumed that they choose the mode and route combination simultaneously such that total operating cost of the network (or system) is minimized as a whole. This assumption permits users to transfer from one mode to another for any given O-D pair. Note that an optimal mix of modes involved is implicitly determined in the traffic assignment process through the integrated network model. A traffic pattern which satisfies this requirement is often referred to as a "system-optimized" pattern, implying that Wardrop’s second equilibrium principle is applied. This system traffic equilibrium may be an idealized target which may not be observed in practice unless cooperation among individual users or a central authority with power to modify individual choices is pre-conditioned. Nevertheless, this can be justifiable at strategic planning level in the sense that the system-optimized flow pattern has the appeal of being the most efficient traffic pattern for society, especially in terms of maximizing utilization of existing capacity. By this assumption, the total cost of the flows of all products in the multimodal network consists of the objective function \( Z_{LDP} \) that we seek to minimize over the set of feasible flows. If we define \( c^k_a \) as the average cost incurred by shipping product \( k \) on link \( a \), this may be expressed as

\[ Z_{LDP} = \sum_{k \in K, a \in A} x^k_a \cdot c^k_a(x_a) \]  

(4)

It is important to note that the objective function must be convex to guarantee a unique, optimal solution. However, the cost functions for freight transportation are not limited to strictly increasing such as those encountered in urban passenger applications. As noted in Friesz, et al. (27), economies of scale may be exhibited for some modes with limited right-of-way such as rail where operating costs initially decline as volume
increases and then begin to increase as capacity limits are approached. When non-increasing cost functions are employed, the objective functions may fail to be convex, giving rise to multiple equilibria. Precise specification of the cost function is a critical problem in network equilibrium models.

The cost function is used to model the cost perceived by users of transportation links and transfers in a freight network and thereby reflects the mode and route choice behavior of network users. In practice, the mode and route chosen by users are not necessarily the fastest and cheapest. Rather the actual mode and route choice reflects users’ perception of the sum of dual cost components, i.e., monetary shipping cost and delivery time. In this respect, the cost we consider is a generalized cost that has these two cost components. To capture the congestion effects on link performance, we additionally assume that the link cost depends on the flow over that link, i.e., \( c^k_a(x_a) \). Then, the generalized link cost function takes the following shape:

\[
 c^k_a(x_a) = SC^k_a(x_a) + \alpha_t \cdot TC^k_a(x_a) 
\]  

(5)

Equation (5) states that the link cost perceived by users is the sum of shipping cost \( SC^k_a \) and monetary value of travel time \( TC^k_a \) incurred by product \( k \) using link \( a \). The coefficient \( \alpha_t \) may vary by commodity type, and can be interpreted as the value of time for product \( k \). It should be non-negative since users are expected to avoid mode and route with longer travel time. The value of the coefficient may be crucial to flow assignment in the MNCP model and therefore should be carefully calibrated for the different commodities under consideration.

Finally, the multimodal network equilibrium model consists of minimizing (4) subject to the constraints (1)-(3). The optimality principle ensures that in the final flow distribution, all paths with positive flows will have the same marginal cost or lower than on the other paths. The problem is referred to as the lower-level problem (LLP) in the MNCP and may be restated in a compact form as follows.

(LLP) Minimize \( Z_{\text{LL}}(x) \), subject to \( x \in H \)

Maximum Flow Problem

Given the optimal distribution pattern of pre-specified freight demand over the network, the problem is now to find a maximum volume of freight that can flow in the network under the prevailing behavioral and physical constraints. Note that the behavioral aspects of network users are already reflected in the aforementioned network optimization model, which produces an equilibrium flow pattern based on system optimal principle of mode and route choices. In contrast, the physical conditions that constrain the flow of freight over the network can be specified in various forms. To facilitate our exposition, we describe each ingredient of the physical constraints in connection with the six core elements identified in the VI-CORE model earlier.

• **Resource Availability:** This condition characterizes the 'Vehicles' component in the VI-CORE model. Obviously total volume of all products moved between an O-D pair \( w \)
cannot exceed the maximum amount of freight that can be served by all the available vehicles, expressed as
\[
\sum_{k \in K} \sum_{m \in M} k_i^m (s) \leq \sum_{m \in M} \delta^m \cdot V^m_w, \quad \forall w \in W
\]  
(6)

where \(\delta^m\) is the average load factor of mode \(m\) measured in tons per vehicle, and \(V^m_w\) is the predetermined fleet size of mode \(m\) available for O-D pair \(w\) in unit time. The value of \(V^m_w\) can be used to examine the system effects of different fleet sizes of each mode.

- **Facility Capacity**: Each facility (i.e., 'Infrastructure') represented as a link on the network has its own capacity, i.e., maximum volume of freight allowed on the link. Thus total freight volume on a link must be within the link’s capacity. To allow delay on link, it may be represented by a link storage capacity proposed by Yang et al. (2000) as below,
\[
\sum_{a \in A} \lambda^a \cdot x^a \leq Q^m_a, \quad \forall a \in A
\]  
(7)

where \(\lambda^a\) is the Lagrange multipliers indicating link queuing delay and \(Q^m_a\) is the maximum storage capacity of link \(a\), measured in tons. Since the links include not only physical line-haul segments of each mode but also goods movements at transfer facilities, these constraints encompass all the physical facilities involved in freight transportation.

- **Maximum Demand Potential**: The maximum demand constraints may correspond to both ‘Customers’ and ‘Externalities’ components of the VI-CORE model in the sense that freight demand results not merely from the interactions between multiple actors, but also from future development of each region in the network. There exists a ceiling of total demand that can be accommodated by each region. To incorporate this condition, we consider the additional demand \((q^a_k)\) distinguished from the existing demand \((q^a)\) and it is assumed as variable. Then the constraints can be described as total volume of product \(k\) produced at an origin region \((r)\) or attracted by a destination region \((s)\) should be within an upper limit of the region for accommodation. These may be represented by
\[
\sum_{a \in A} \sum_{a \in A} q^a_k \leq O^m_{r}, \quad \forall r \in R \subseteq N, k \in K
\]  
(8)
\[
\sum_{a \in A} \sum_{a \in A} q^a_k \leq D^m_s, \quad \forall s \in S \subseteq N, k \in K
\]  
(9)

where \(O^m_{r}\) and \(D^m_s\) are a preset upper bound of product \(k\) produced at origin \(r\) and attracted by destination \(s\), respectively.

- **Level of Service**: The level-of-service (LOS) constraints may reflect the ‘Operations’ components in the VI-CORE model. The maximum flow bound in the network corresponds to the varied level of service quality provided in the network. Considering the growing importance of delivery time in the logistics community, we use average travel time as a LOS measure. That is, the average travel time required for product \(k\) to be transported between O-D pair \(w\) (\(t^w_k\)) should be less than the maximum allowable travel time for the O-D pair (\(T^w_{max}\)), represented by
\( l^w_k(s) \leq \delta_k \cdot T^w_{\text{max}}, \quad \forall k \in K, w \in W \)  \hspace{1cm} (10)

where \( \delta_k \) is a commodity-specific parameter ranging from 0 to 1. \( T^w_{\text{max}} \) is a preset upper bound of travel time for O-D pair \( w \) and may vary by commodity-type. As a result, the LOS constraints ensure a minimum LOS in the network while considering the value of travel time of each product.

- **External Cost:** External costs incurred by accidents and air pollution may also constrain the maximum freight volume allowed on the network, thereby associated with the 'Regulations' and 'Externalities' components in the VI-CORE model. Using the monetary cost of various external factors estimated for freight transportation such as in Forkenbrock (27), we can symbolize the external cost constraints as follow:

\[
\sum_{e \in E} \phi_e \cdot x_e \cdot d_{e,a} \leq EC^w_{e,a} \quad \forall a \in A
\]  \hspace{1cm} (11)

where \( \phi_e \) is the unit cost per ton-mile of external factor \( e \in E \) where \( E \) is the set of external factors, \( d_{e,a} \) is the distance of link \( a \) measured in mile, and \( EC^w_{e,a} \) is the maximum allowable external cost per unit time on the link \( a \).

- **Flow Conservation and Non-negativity:** As usual, a set of flow conservation and non-negativity constraints are imposed to obtain the physically meaningful solution set.

\[
\sum_{s \in S_{\text{in}}} x_{n,s} - \sum_{s \in S_{\text{out}}} x_{n,s} = \begin{cases} 
q & \text{for } n = s \\
- q & \text{for } n = r \\
0 & \text{for all } n \in N - \{r,s\}
\end{cases} \quad \forall n \in N
\]  \hspace{1cm} (12)

\[
x_{n,s} \geq 0, \quad \forall k \in K, \quad r,s \in N
\]  \hspace{1cm} (13)

where \( I_n, O_n \subseteq A \) represent the set of links entering and leaving node \( n \), respectively.

Note that the link flow pattern \( x \) is determined by solving the LLP. Again, to determine a unique solution of the problem, an optimal criterion must be established over the set of feasible solutions satisfying the constraints (6)-(14). Under the equilibrium distribution of existing demand, maximum flow in the network depends on the amount of additional demand allowed on the network. This implies that the target to be maximized in the problem should be the sum of additional demand of each product accommodated by the network for all O-D pairs. That is

\[
\text{Maximize } Z_{\text{UPL}} = \sum_{k \in K} \sum_{w \in W} q^k_w
\]  \hspace{1cm} (14)

Consequently, the maximum flow problem consists of maximizing (14) subject to the constraints (6)-(13) and the LLP. The problem is referred to as the upper-level problem (ULP) in the MNCP and becomes the master problem since it contains main objective of the MNCP. If we denote by \( F \) the set of feasible solutions of the ULP and by \( q \) the vector with elements \( q^k_w \), its compact mathematical form may be represented as

\[
\text{(ULP)} \quad \text{Maximize } Z_{\text{UPL}}(q), \quad \text{subject to } q \in F
\]
Bi-Level Model

The basic concept of the bi-level approach for the MNCP essentially involves solving two optimization problems, i.e., the maximum flow problem and the integrated network equilibrium problem. As illustrated in Figure 2, these are complementary because they capture different aspects of the network: The maximum flow problem models network capacity by maximizing the flow of freight on the network while maintaining an equilibrium network condition. On the other hand, the integrated network equilibrium problem characterizes the mode and route choice behavior of network users in terms of a minimum system cost. Taken together, they comprise the basic ingredients of the MNCP.

FIGURE 2. Conceptual Framework of Bi-Level Optimization Model

It is important to note that an optimal value of network capacity may be different from various perspectives. Thus substantial structural differences of the models used to estimate network capacity can unfold, depending upon the focus and scope of the study for capacity analysis. Our study was intended to develop a versatile model that can be applicable to the network capacity problems involving combinations of modes as well as individual modes. Ideally, capacity models should be examined within a long-term time framework where shifts in productions and consumptions are likely to occur in many regions over the network. The issue is how to select the target (future) O-D matrix appropriately. The target O-D demand should correspond to the future development and growth of each region. In this respect, the MNCP based on a fixed demand pattern can be extended further to incorporate the variable demand situations in which destinations chosen by users can vary by network conditions. In this context, the LLIP can be replaced by a combined trip distribution, assignment and modal split model such as the one developed by Friesz (23).

When integrating any combined model with the MNCP, it is also noteworthy that traditional demand models dealing with destination choice of network users should be altered to freight transportation. As is well known, freight demand is derived from economic activities of multiple actors in the economic system. Thus its distribution is highly dependent on the characteristics of the regions where commodities transported are ultimately consumed. This is in contrast to that of passenger transportation in which trip generation potential at the origin regions is generally regarded as most important. From this economic perspective, freight distribution depends on the choice of a source of
supply among many potential regions that can provide commodities required at the consumption region. This implies that existing models based on the destination choice users should be revised to reflect the distinct features of freight transportation. In this setting, the maximum flow on the network may be translated as the maximum demand of all products accommodated by all destinations in the network.

CONCLUSIONS AND FUTURE RESEARCH

In this paper, we proposed a logical network model based on bi-level programming that can be used to evaluate the capacity of multimodal freight transportation systems. Capacity assessment is a structured and analytical process whereby the various dimensions of capacity must be considered within the broader systems context, as well as evaluated for specific entities within the systems. If some important dimensions of capacity are missed, then the chances of successfully securing sustainable capacities are diminished. Therefore, capacity assessment and development must go beyond the level of the individual elements and the entities to ensure that capacities at all levels are both addressed and properly utilized. The XNCP model devised from the systems perspective appears comprehensive in the sense that many key factors crucial are incorporated, including multiple modes and products, the behavioral aspects of network users, external factors as well as the physical characteristics of network. The model will help us facilitate efficient use of existing capacity and also to plan for new capacity which is most appropriate from social, economic and environmental perspectives.

We should point out that the capacity model we proposed is at the early stage of development and still incomplete. The reasonableness and applicability of the developed model must be tested and assessed in the next stage. This will be performed through an application to real-world capacity problems in order to ensure that the model is practical and applicable. Current capacity problems developing in the Southern California area should provide an ideal case study for the application of our model. The development of efficient algorithms will be essential to solve the large-scale XNCP. Based on the results of the case study, potential solution approaches to capacity-related questions will be described and discussed. Specific issues or obstacles that might hinder the multimodal system capacity analysis will be also addressed in this phase.

ACKNOWLEDGEMENT

This research was partially supported by the University of California Transportation Center (UCTC) and the California Department of Transportation (Caltrans). The authors gratefully acknowledge that support. Any opinions expressed are those of the authors and not necessarily those of the UCTC or Caltrans.

REFERENCES


