A Network Capacity Model for Multimodal Freight Transportation Systems

Minyoung Park*· Yong Jin Kim**

목 차

Key Words : Transportation Capacity, Freight Transportation Planning, Multimodal Transportation System

Abstract

This paper presents a network capacity model that can be used as an analytical tool for strategic planning and resource allocation for multimodal transportation systems. In the context of freight transportation, the multimodal network capacity problem (MNCP) is formulated as a mathematical model of nonlinear bi-level optimization problem. Given network configuration and freight demand for multiple origin-destination pairs, the MNCP model is designed to determine the maximum flow that the network can accommodate. To solve the MNCP, a heuristic solution algorithm is developed on the basis of a linear approximation method. A hypothetical exercise shows that the MNCP model and solution algorithm can be successfully implemented and applied to not only estimate the capacity of multimodal network, but also to identify the capacity gaps over all individual facilities in the network, including intermodal facilities. Transportation agencies and planners would benefit from the MNCP model in identifying investment priorities and thus developing sustainable transportation systems in a manner that considers all feasible modes as well as low-cost capacity improvements.

* Principal investigator, Post Doctoral Researcher, Institute of Transportation Studies, University of California, Irvine, mypark@uci.edu,

** Corresponding author, Assistant Professor, Asia Pacific School of Logistics, Inha University, yongjin@inha.ac.kr, (032) 860-8233,

I. Introduction

Capacity assessment and development are integral to the management of people, organizations, assets and the broader systems within which they function. The term capacity is generally defined as the ability of an entity (individual, organization or facility) and a system (a set of entities) to perform its function or to produce products. In the context of transportation, capacity particularly plays a key role as performance measure in the transportation system planning and investment decision-making process. Overestimation of capacity can result in poor system performance, while underestimation of capacity can lead to unnecessary investments in infrastructure. Both cases eventually yield undesirable outcomes in the management of transportation systems, either suffering from inadequate capacity to meet anticipated transportation demand or causing inefficient use of resources. Thus, the incorporation of precise capacity measure into the planning process is essential for the development of sustainable transportation systems and better accountability of investment decisions.

Due to the importance of capacity assessment in transportation planning, capacity modeling for transportation systems has been the subject of intense study over the past few decades. As public policies for transportation planning changed during the 1990s toward a modally balanced transportation system, transportation researchers and planners have begun to pay more attention to the development of analytical tools and data to implement a multimodal transportation planning process that considers not only all feasible modes but also low-cost capacity improvements (TRB, 1999; Kale, 2003). The changes from modal fragmentation to cross-modal coordination and from system construction to system optimization require transportation planners to more precisely identify and measure capacity gaps across all transportation facilities. However, capacity models developed to date for use in transportation appear insufficient for multimodal systems capacity analysis.

Most previous state-of-art capacity analyses have focused on estimating the capacity of individual transportation facilities. A comprehensive review of capacity models falling in this category may be found in a recent NCHRP report (Cambridge Systematics, Inc. et al., 1998), in which the Highway Capacity Manual

(TRB, 1994) serves as a good example. Capacity assessment is a structured and analytical process in which the various dimensions of capacity need to be assessed within the broader systems context, as well as evaluated for specific entities within the systems. If some important dimensions of capacity are overlooked, then the chances of successfully securing sustainable capacities are diminished. Hence, capacity assessment and development must go beyond the level of the individual entities to ensure that capacities at all levels are both addressed and properly utilized and sustained.

Only a few examples are found in the literature addressing the inter-relationship between simultaneous consideration of various system elements and system-wide capacity. But these are limited to either measuring the capacity of a corridor (Cambridge Systematics, Inc. et al., 1998; TRB, 2000), or capturing the capacity of a single-mode network (Morlok and Riddle, 1999; Yang et al., 2000; Ge et al., 2003). As pointed out by Park and Regan (2005), even the system approach to capacity assessment would suffer from some shortcomings when evaluating the capacity of a multimodal system. A main drawback of this approach lies in the fact that it does not account for the likely impacts of intermodal substitution and complementarity on system capacity. Furthermore, it fails to take the existence of multiple actors with different objectives into consideration. Therefore, the capacity information based on the existing system capacity concept is of only limited value.

In fact, the failure of many projects and programs for capacity improvement can often be attributed to the narrow view of capacity that had been used. From a multimodal perspective, capacity problems are a function of inadequate consideration of broader system factors, poor integration and coordination of various system dimensions, and/or dependency on external factors such as land use, environment, and technology. This implies that capacity problems facing certain modes may be overcome by enhancing the utilization of residual capacity in other modes. This is especially true in the multimodal systems for freight transportation, which involve the use of multiple modes and facilities as well as the compound interactions of multiple actors. Even though simultaneous consideration and integration of a wide range of system components trigger greater complexity in capacity modeling, new emphasis needs to be on a much broader systems approach to capacity analysis.

Based on a bi-level programming approach, we recently proposed an analytical framework for the capacity assessment of multimodal transportation networks (Park and Regan, 2005). As an extension of the previous work, the substantive structure of the conceptual model is unfolded in this paper in the form of a mathematical model, with its solution algorithm and application to a test network. New information and advanced methods for capacity analysis would help transportation agencies and planners identify investment priorities across all feasible transportation modes, leading to better allocation of resources and more efficient utilization of existing capacity in the transportation system planning and management.

The paper is organized in the following way. The network representation chosen to integrate multiple modes and intermodal movements is first briefly described. The multimodal network capacity problem is then formulated as a bi-level optimization problem. This is followed by the description of a heuristic solution algorithm adopted to solve the problem. A numerical example is provided to illustrate the application of the model and the effectiveness of the solution algorithm. Finally, the paper concludes by discussing future research directions.

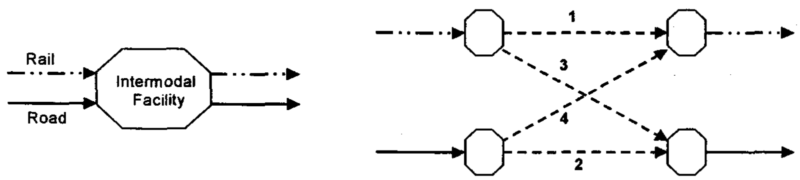
II. Representation of Multimodal Network

The modeling framework of a Multimodal Network Capacity Problem (MNCP) identified in the earlier work is essentially that of a multimodal freight network made up of various facilities, on which multiple products are transported by either one mode or the combined use of different modes between given origin and destination points. For arranging information about the characteristics of network and freight movements over the network, we introduced an integrated network in which each mode has its own distinct physical network and these modal networks are integrated through intermodal facilities.

It should be noted that the MNCP model was designed to serve as an analytical tool that can be used within a strategic level of planning for the development of transportation investment strategies. As indicated by Guelat, Florian and Crainic (1990), the strategic level of planning implies a long-term horizon and deals with a relatively large geographic area in the scope of analysis. Thus the level of detail for network representation should be appropriate for strategic planning.

The integrated network can be represented by a graph that includes a set of nodes and a set of directed arcs (or links). The set of nodes characterizes all product origins and destinations, rail yards and stations, air or marine ports, intermodal transfer facilities and the intersections of different line-haul segments. The set of links represents line-haul segments that connect two different nodes, such as roads and rail lines. Each link has its natural attributes, including length, vehicle types allowed, and capacity. Freight volume and shipping costs are also associated with each link. Parallel links are allowed to model goods movements by different modes available between two adjacent nodes. Assigning a separate link to each mode enables us to easily identify not only the flow of goods by different modes on the same route, but also different types of services and different carriers in the same mode.

Mode to 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 costs and delays. This can be achieved 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. An example of expanding physical network is presented in Figure 1, in which conceptual links 1 and 2 represent transshipment between different vehicles in the same mode while links 3 and 4 characterize transfer movements between different modes. In a similar way, the conceptual links may also be used to allow the modeling of other terminal activities such as loading, unloading and consolidation, as well as pick-up and delivery movements at each origin and destination, respectively.



(a) Physical network (b) Expanded network

<Figure 1> Representation of intermodal transfer movements.

This type of network representation has been successfully implemented with small variations in other contexts, for example, a multimode multiproduct network assignment problem (Guelat et al., 1990) and a multimodal network design problem

(Loureiro and Ralston, 1996). In modeling a multimodal network, the explicit inclusion of transfer links as well as parallel links would be useful especially in terms of visualizing all possible combinations of freight movement over the network. In this context, intermodal transfer movements can be defined as an additional means of transportation available in the network that has its own unique characteristics, such as capacity and cost functions.

III. The Multimodal Network Capacity Model

Once a multimodal network is defined, the problem at hand is to find the maximum flow that the network can accommodate. As discussed, there exist in multimodal systems different levels of decision-makers, each with different objectives. In general, decision-makers at the higher levels solve problems while taking into account the responses at the lower levels in the form of constraints. In these cases, an optimal solution of the higher level problems can be achieved after the lower level problems have reached optimal (or equilibrium) conditions. This is especially true of the MNCP, in which the identification of capacity gaps over a multimodal network is the ultimate goal and this can be attained only after knowing the level of current capacity utilization by existing demand. Problems of this form are known in the literature as bi-level programming problems. These provide a basis in developing a mathematical form of the MNCP model presented in this section.

1. Basic Assumptions

In the context of strategic planning, the demand for freight transportation services is often generated from national freight flow statistics or economic input-output models. Thus freight demand is assumed to be given and exogenous to the model. The demand is generally categorized by different classes of products, each of which with distinct demand characteristics and different valuation for the cost factors. It takes a form of origin-destination matrices corresponding to the set of commodities under consideration, measured in tons per unit time. For simplicity of analysis, only commodities that can be shipped by all modes available in the network are considered because these are appropriate for intermodal transfers.

The level of aggregation for strategic planning also leads to the assumption that shippers and carriers are not considered as distinct actors in the decision-making process of mode and route choices for shipping freight. Instead, it is assumed that they make their mode and route choice decisions simultaneously in a collaborative manner such that goods are shipped at a minimum cost. Mode choices made by shippers are thus modeled in combination with route choices made by carriers. Note that the model explicitly considers competition among the modes, although it does not account for a variety of micro-circumstances where individual shippers'decisions are regarded as a significant factor. Shipments are assigned to the paths of least cost according to the user equilibrium (UE) principle in which at user equilibrium, the travel cost on all used paths is equal, and also less than or equal to the cost on any unused path.

Path cost is calculated as a function of link costs and travel times. The fixed cost of a physical link is dependent on and proportional to its length and the unit cost per ton-km of each mode being used. On the other hand, link travel time is a function of the length of the link and the average speed of the mode. In order to capture the congestion effects in the network, it is assumed that link travel times are dependent on link flows for shipments assigned to each link, except for rail links where link travel time is constant and thus free of congestion. These functions are calibrated as polynomial-shaped, nonlinear delay functions with volume and capacity values expressed in tons per unit time. A traffic assignment algorithm is used to distribute shipments over the multimodal network according to the aforementioned UE criterion.

2. Mathematical Formulation

Based on the bi-level approach and assumptions described above, a general model for estimating the capacity of a multimodal transportation network can be formulated in the following way. Several constraints associated with physical and operational conditions of the network can be specified in various forms and embedded in the MNCP model. These are basically taken from the study of Park and Regan (2005), which identifies six key factors that constrain the flow of freight in the multimodal network, and also explains the basic relationship between each factor and system capacity in detail.

Let $G(N, A)$ define a multimodal network that consists of a set of nodes, $n \in N$ and a set of directed arcs (links), $a \in A$. A subset of A, denoted by A_m includes the links allowing a mode $m \in M$. To define an origin-destination (O-D) pair, we distinguish two special nodes in G: an origin node $r \in R \subseteq N$ and a destination node $s \in S \subseteq N$. Note that the origin and destination node sets are not mutually exclusive since the nodes can serve simultaneously as origins and destinations for different shipments (i.e., $R \cap S \neq \varnothing$). Note also that all nodes are not required to be origins and destinations. Each O-D pair r-s is connected by a set of paths (routes) through the network, denoted by $k \in K_{rs}$. The following notations are used throughout the paper:

q_{rs} = total demand on O-D pair r-s, in tons;

x_a = flow volume on link $a \in A$, in tons;

$t_a(x_a)$ = travel time on link $a \in A$ as a function of link volume x_a, in hours;

$c_a(x_a)$ = travel impedance on link a as a function of link volume x_a, in dollars per ton;

f_k^{rs} = total flow on path k connecting O-D pair r-s, in tons;

C_k^{rs} = travel impedance on path k connecting O-D pair r-s, in dollars per ton;

T_k^{rs} = travel time on path k for O-D pair r-s, where $T_k^{rs} = \sum_{a \in A} \delta_{a,k}^{rs} \cdot t_a$, in hours;

$\delta_{a,k}^{rs}$ = indicator variable, 1 if link a is on path k between O-D pair r-s, 0 otherwise.

Then, the MNCP model is formulated as the following bi-level programming problem:

Maximize $$Z_{ULP} = \sum_{r \in R} \sum_{s \in S} q_{rs} \tag{1}$$

subject to

$$x_a(\mathbf{f}) \leq Q_a, \qquad \forall a \in A \tag{2}$$

$$\sum_{a \in A_m} \phi_m \cdot x_a(\mathbf{f}) \cdot t_a(x_a(\mathbf{f})) \leq V_m \cdot H_m, \qquad \forall m \in M \tag{3}$$

$$T_k^{rs} = \sum_{a \in A} \delta_{a,k}^{rs} \cdot t_a(x_a(\mathbf{f})) \leq T_{rs}^{\max}, \qquad \forall r \in R,\ s \in S,\ k \in K \tag{4}$$

$$O_r = \sum_{s \in S} q_{rs}(\mathbf{f}) \leq O_r^{\max} - \bar{O}_r, \qquad \forall r \in R \tag{5}$$

$$D_s = \sum_{r \in R} q_{rs}(\mathbf{f}) \leq D_s^{\max} - \bar{D}_s, \qquad \forall s \in S \tag{6}$$

$$\sum_{m \in M} \sum_{a \in A_m} \gamma_e^m \cdot x_a(\mathbf{f}) \cdot d_a \leq EF_e^{\max}, \qquad \forall e \in E \tag{7}$$

$$\sum_{a \in I_n} x_a(\mathbf{f}) = \sum_{a \in O_n} x_a(\mathbf{f}), \qquad \forall n \in N - \{r, s\} \tag{8}$$

$$x_a, q_{rs} \geq 0, \qquad \forall a, r, s \tag{9}$$

and the set of path flows $\mathbf{f} = (\cdots, f_k^{rs}, \cdots)$ satisfies the user equilibrium conditions, which can be obtained by solving the following equivalent optimization problem:

$$\text{Minimize} \qquad Z_{LLP} = \sum_{a \in A} \int_0^{x_a} c_a(x)\, dx \tag{10}$$

subject to

$$\sum_{k \in K_{rs}} f_k^{rs} = q_{rs}, \qquad \forall r \in R,\ s \in S \tag{11}$$

$$f_k^{rs} \geq 0, \qquad \forall r \in R,\ s \in S,\ k \in K_{rs} \tag{12}$$

and the link flow relationship

$$x_a = \sum_{r \in R} \sum_{s \in S} \sum_{k \in K_{rs}} \delta_{a,k}^{rs} \cdot f_k^{rs}, \qquad \forall\, a \in A \tag{13}$$

where

Q_a = practical capacity of link a, in tons;

ϕ_m = conversion factor from freight volume to vehicle, in vehicles per ton;

V_m = total number of vehicles available in mode m, in vehicles per unit time;

H_m = average operating hours of vehicles in mode m, in hours;

$T_{rs}^{\max}$ = maximum allowable travel time between O-D pair r-s, in hours;

$O_r^{\max}$, $D_s^{\max}$ = maximum potential demand produced at origin r and consumed at destination s, respectively, in tons;

$\bar{O}_r$, $\bar{D}_s$ = existing demand produced at origin r and consumed at destination s,respectively, in tons;

γ_e^m = unit rate of external factor $e \in E$ produced by goods movement in mode m , where E is the set of external factors, in unit rates per ton-km;

d_a = the distance of link a, in kilometers;

$EF_e^{\max}$ = maximum allowable rate of external factor e, in unit rates per ton-km;

$I_n, O_n \subseteq A$ = the set of links entering and leaving node n, respectively.

The objective of the upper level problem (Z_{ULP}) is to maximize the sum of the freight volumes that can be accommodated over the elements of the sets R and S in the network. The resultant value is the quantity that can be interpreted as network capacity. Inequality (2) represents a facility capacity constraint, ensuring that flow on each link cannot exceed the capacity of the link. Constraint (3)characterizes the limitation of fleet operations available in each mode. Obviously, the total volume of freight carried by a mode is limited to an upper bound of the number of vehicles available in the mode during the period under consideration. Constraint (4)is a level of service constraint guaranteeing that the average travel time required to transport products between an O-D pair must be less than the maximum allowable travel time for the O-D pair. The maximum flow in the network often depends on the level of service provided by the network. As delivery time is increasingly regarded as an important factor in the logistics community, the inclusion of this constraint is essential in terms of maintaining a minimum level of service in the network.

Constraints (5) and (6) represent the maximum demand potential in the network. Freight demand results not only from the interactions between multiple actors, but also from the future development of each region and economic activities at the region. However, the amounts of freight volume produced and consumed at each region should be within an upper limit of the region for accommodation. These constrain the flow of goods in the network, and thus must be incorporated as constraints. On the other hand, constraint (7)represents the externality constraints stating that total production rate of an external factor incurred by freight transportation, such as accidents or emissions must be less than a preset upper bound of the external factor in the network. A set of flow conservation constraints in (8)and non-negativity constraints for decision variables in (9) are also required to ensure that the solution of the problem be physically meaningful.

Finally, constraints (10)-(13) represent the integrated multimodal network equilibrium problem that finds the set of link flows that satisfy the UE criterion when all the OD demand has been properly assigned. As demonstrated in Sheffi (1985), the first-order conditions for the minimization problem is identical to the UE conditions and thus an equilibrium flow pattern is obtained by solving the problem. In the network equilibrium problem, the objective function is the sum of the integrals of the link performance functions. Equation (11)simply states the conservation of flow meaning that the flow on all paths connecting each O-D pair has to equal the demand for the O-D pair. Non-negativity constraint in (12)is also imposed on the decision variables, while link flow and path flow are related in equation (13).

As presented, the bi-level programming approach for the MNCP essentially involves solving two optimization problems, i.e., Maximum Flow Problem (MFP) and integrated network equilibrium problem. These are complementary because they capture different aspects of the network. The MFP models network capacity by maximizing the flow of freight on the network while satisfying a set of network constraints including an equilibrium flow pattern. In contrast, the integrated network equilibrium problem characterizes an equilibrium flow pattern that reflects the mode and route choice behaviors of network users in terms of a minimum cost. Taken together, they comprise the basic ingredients of the MNCP model.

3. Cost Function and Uniqueness

It is important to note that the link cost function in the objective function of the lower level network assignment problem as in equation (10)captures the cost perceived by users of the link. This cost includes the line haul cost and any transfers incurred using the link and thereby reflects the mode and route choice behavior of network users. In practice, the mode and route chosen by users are not necessarily either the fastest or cheapest. Rather the actual mode and route choice reflects users'perception of the sum of the dual cost components, i.e., monetary shipping cost and delivery time. In this respect, the cost that we consider is a generalized cost that includes these two cost components. To reflect the congestion effects on the link performance function, we assume that the link cost depends on the flow over that link, i.e., $c_a = c_a(x_a)$. Then, the generalized link cost function takes the following shape:

$$c_a(x_a) = SC_a + \theta \cdot t_a(x_a) \qquad (14)$$

where

SC_a	= shipping cost on link a, where $SC_a = r_a \cdot d_a$, in dollars per ton;
r_a	= base shipping rate for link a, in dollars per ton-km;
θ	= average value of time of products, in dollars per hour;
$t_a(x_a)$	= travel time on link a, where $t_a(x_a) = t_a^0 \cdot [1 + \alpha \cdot (x_a / Q_a)^\beta]$, in hours;
t_a^0	= free-flow travel time on link a, where $t_a^0 = d_a / s_a$, in hours;
s_a	= average free-flow speed on link a, in km/hr;
α, β	= coefficients in the polynomial delay function.

The link impedance function in equation (14)consists of the sum of shipping cost and monetary value of travel time incurred by using link a. The coefficient θ may vary by commodity type, being interpreted as the average value of time for products transported in the network. It should be non-negative since users are expected to avoid modes and routes with longer travel times. The value of the coefficient may be crucial in the network assignment process, and thus should be carefully calibrated for the commodities under consideration.

It is also important to note that the objective function must be convex to guarantee a unique, optimal solution. As mentioned before, the shipping cost of a link is proportional only to its length, thus increasing as the link length increases. In addition, link travel time is a function of link volume, and has the polynomial-type, nonlinear structure of the function due to the congestion effects in the network as shown in the equation (14). This implies that the link cost functions considered are continuous and monotonically increasing with the amount of link flow, consequently ensuring a unique flow pattern.

IV. Solution Procedure

The mathematical model of the MNCP takes a form of nonlinear bi-level optimization problem with linear objective function subject to nonlinear constraints. The solution of the maximum flow problem in the upper level is conditional on a set of user optimal flows assigned to different paths serving each O-D demand in the network. The UE flow pattern can be determined by solving the lower level network assignment problem. A heuristic algorithm was developed to solve the bi-level problem, based on a linear approximation technique combined with a standard traffic assignment algorithm as presented below.

1. Linear Approximation of the Maximum Flow Problem

The maximum flow problem embedded in the MNCP model has a simple linear objective function, but includes nonlinear constraints. For solving this type of problem, an iterative, linear approximation method has been proposed and demonstrated to be efficient (Yang et al., 2000). As in equations (3) and (4), the nonlinear constraints are associated with link travel time, which can be approximated to be linear at major iteration n using the following first-order Taylor's expansion.

$$t_a^{(n)}(x_a) \approx t_a(x_a^{(n-1)}) + \left.\frac{\partial t_a}{\partial x_a}\right|_{x_a = x_a^{(n-1)}} \cdot (x_a - x_a^{(n-1)}),\ \forall a \in A \tag{15}$$

where x_a^{n-1} is the link flow obtained from the last iteration $n-1$. Based on the linear approximation (LA) technique, the MFP is simplified to a linear programming (LP) problem that can be easily solved using the standard simplex method. The solution of the approximate LP problem will generate a new solution set to the original bi-level problem, i.e., the maximum volume of O-D flow that can be accommodated by the feasible paths over the O-D pairs in the network.

2. UE Network Assignment Algorithm

In finding feasible paths that can accommodate additional flows on each O-D pair, a path construction model demands as input link impedance values, which can be calculated using the freight volumes that resulted from the last iteration. This implies that the lower level UE assignment problem formulated in equations (11)-(13)needs to be solved to create the feasible paths at each iteration. The well-known convex combination method, also known as Frank-Wolfe algorithm in the literature (Sheffi, 1985), can be used to solve the lower level problem as follows:

Step 0: Initialization. Perform all-or-nothing assignment for all commodities $p \in P$ based on the set of free-flow link impedance values $\{c_a(0)\}$ and determine an initial set of link volumes $\{x_a^0\}$. Set $n = 0$

Step 1: Cost update. Calculate link cost $c_a^n = c_a(x_a^n)$, $\forall a$

Step 2: Direction finding. Perform another all-or-nothing assignment based on c_a^n and obtain a set of auxiliary flows $\{y_a^n\}$

Step 3: Line Search. Find α^n by solving $\min_{0 \le \alpha \le 1} \sum_a \int_0^{x_a^n + \alpha(y_a^n - x_a^n)} c_a(w)dw$

Step 4: Move. Set $x_a^{n+1} = x_a^n + \alpha^n (y_a^n - x_a^n)$, $\forall a$

Step 5: Convergence test. Stop if the following convergence criterion is met, $|x_a^{n+1} - x_a^n| \le \varepsilon$ for all a, where ε is a preset tolerance; otherwise, set $n = n+1$ and go to Step 1.

3. Overall Solution Procedure for the MNCP Model

Based on the methods explained above, the whole bi-level multimodal network capacity problem can be solved by iterating the lower level UE assignment problem and the upper level maximum flow problem, as illustrated in Figure 2. The solution algorithm is described below:

Step 0: Initialization. Given the network configuration, the algorithm is started with an initial set of O-D demand $\{q_{rs}^{(0)}\}$. Set iteration count $n=0$.

Step 1: UE assignment. Solve the assignment problem for the demand set $\{q_{rs}^{(0)}\}$ using the convex combination method, and get a set of UE assigned link flows $\{x_a^{(n)}\}$.

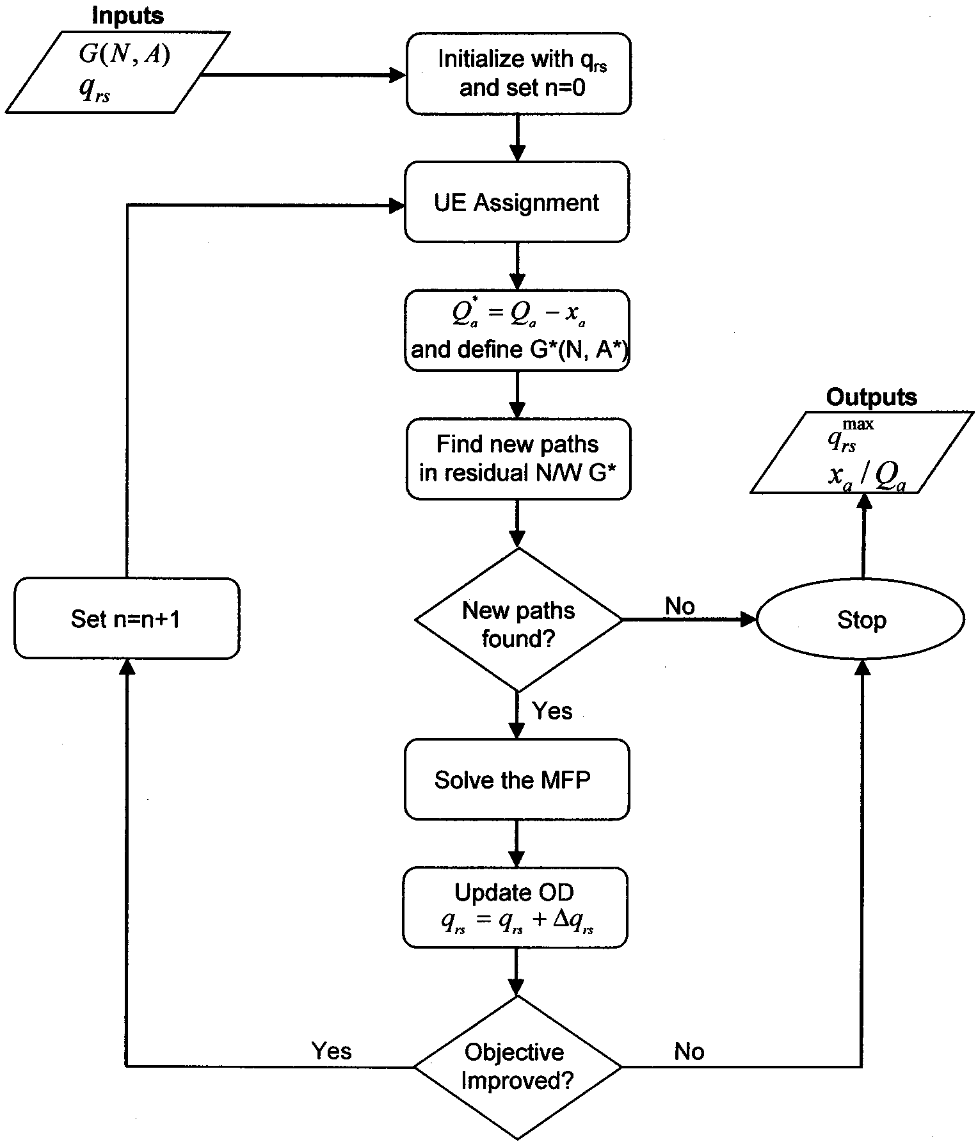
Step 2: Residual network. Calculate the reserve capacity of each link by subtracting the assigned link flow from the link capacity, i.e., $Q_a^{*(n)} = Q_a^{(n)} - x_a^{(n)}$, where $Q_a^{*(n)}$ is the residual capacity of link a at iteration n. Then, define the residual network as $G^{*(n)}(N, A^*)$ consisting of links that can admit more flow up to $Q_a^{*(n)}$.

Step 3: Path construction. Find new paths in $G^{*(n)}$ that can accommodate additional flows between each OD pair. If no paths are found, then stop. Otherwise, go to next step.

Step 4: Maximum flow problem. Solve the MFP using the LA method and obtain a set of additional path flows $\{f_k^{rs(n)}\}$ for each O-D pair r-s.

Step 5: Update OD demand. Update O-D demand for each O-D pair using the equation $q_{rs}^{(n+1)} = q_{rs}^{(n)} + \Delta q_{rs}, \forall r,s$, where $\Delta q_{rs} = \sum_k f_k^{rs(n)}$

Step 6: Convergence test. If $\left| q_{rs}^{(n+1)} - q_{rs}^{(n)} \right| \le k$ for all O-D pairs then stop where k is a preset tolerance. Otherwise set $n = n+1$ and return to Step 1.



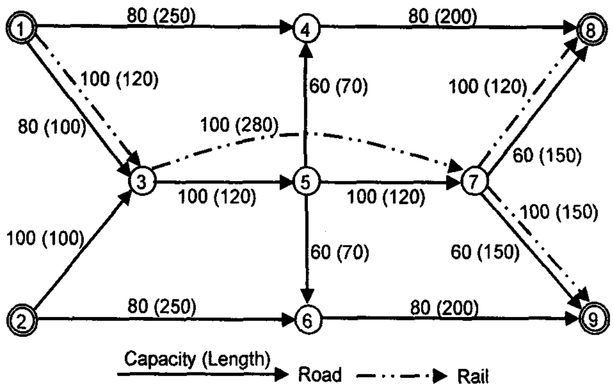
<Figure 2> Overall solution procedure for the MNCP model.

V. A Numerical Example

To illustrate the application of the MNCP model and to evaluate the effectiveness of the solution algorithm presented in the earlier section, a hypothetical exercise is performed and described in this section.

1. Problem Setting

A multimodal test network was developed for this hypothetical exercise. As shown in Figure 3-(a), the network consists of 9 nodes and 16 directed physical links. Only two modes, i.e., on-road truck and rail are considered in this example, and these networks are integrated through two intermodal facilities located at nodes 3 and 7, respectively. To represent intermodal transfer movements, these nodes are expanded in a manner as in Figure 1. The resultant expanded network is composed of 15 nodes and 24 links. The numbers close to each link in the network indicate the link's practical capacity, while the numbers in parentheses are the lengths of the links. The capacity of each intermodal facility is assumed to be 5,000 tons per month. Loading and unloading at the origin and destination and any other types of terminal activities are not considered in this exercise.



(a) Test network

O/D	8	9	O_r	O_r^{max}
1	50	80	130	180
2	70	50	120	170
D_s	120	130	250	
D_s^{max}	180	200		

Measured in 1,000 tons per month

(b) Existing O-D matrix and zonal demand potential

<Figure 3> The test network and associated freight demand.

The test problem to be solved by the MNCP model assumes the analysis of a single commodity with the value of travel time of $1.00 per ton-km. Figure 3-(b) shows the existing origin-destination freight demand and maximum demand potential at each zone, measured in 1,000 tons per month, to be distributed over the test network with nodes 1 and 2 representing the origins and nodes 8 and 9 representing destination points.

Link travel times are calculated according to a polynomial delay function of the well known BPR (Bureau of Public Roads) type with coefficients α and β equal to 0.15 and 4.0, respectively. In calculating link impedance function, the capacity of all rail links is set to a very high value to reflect the fact that those links generally operate based on a fixed schedule and thus are free of congestion effects. In addition, the shipping rates relating to the commodity are assumed to be $0.04 per ton-km on road and $0.02 per ton-km on rail. A fixed unit transfer cost of $1.00 per ton is assigned to transshipment operations at both intermodal facilities. It is also assumed that all shipments must be delivered within 72 hours (i.e., 3 days) in the network and all vehicles are available for 24 hours a day. Table 1 summarizes the modal attributes adopted for this empirical analysis. It should be noted that externality constraints are not considered in this exercise due to the difficulty in acquiring reliable data on the unit costs on each external factor. Also note that all the parameters presented here were selected solely for the purpose of this hypothetical exercise without any calibration with real data.

<Table 1> Modal Attributes of the Test Network

Mode	Rate ($/ton-km)	Free-flow speed (km/hr)	Number of vehicles	Load factor (tons/veh)
Road	0.04	60	2,000	10
Rail	0.02	30	500	200
Transfer	1.00	-	150	10

2. Analysis Results and Policy Implications

The problem described above was solved using the solution algorithm proposed for the MNCP model to estimate the capacity of the multimodal test network. As presented in Table 2, the analysis results show that the network can handle up to 322,000 tons of freight volume during the time period under consideration. Compared to current freight volume moved in the network, it represents approximately 30 percent increase.

<Table 2> Base Flow Rate, Maximum Allowable Flow Rate and Reserve Capacity Estimated for O-D Pairs and Network (1,000 tons per month)

O-D pair	Base flow rate (q_{rs})	Maximum flow rate (q_{rs}^{max})	Reserve Capacity ($q_{rs}^{*} = q_{rs}^{max} - q_{rs}$)
1-8	50	60	10
1-9	80	120	40
2-8	70	76	6
2-9	50	66	16
Total	250	322	72

In addition to the network-wide capacity, the MNCP model can also estimate the maximum flows that can be accommodated on all paths for each origin and destination pair. As shown in Table 2, it is estimated that the test network can handle up to 120,000 tons of freight volume between origin 1 and destination 9, indicating that the network still has room for accommodating additional volume of freight on the O-D pair 1-9 up to 40,000 tons. This is comparable to the reserve capacity of the O-D pair 2-8 which is limited to only 6,000 tons and thus can be identified as the most critical corridor.

It is important to note that the sum of these volumes is one measure of residual network capacity. By comparing with the given maximum demand potential of each zone, the predicted maximum flow and reserve capacity estimated for each O-D pair can be used to estimate to what extent zonal economic growth could be accommodated by the existing transportation facilities. Thus, these results may provide some policy implications for the development of sustainable transportation systems associated with future zonal land use and economic growth.

Another important outcome that can be obtained from the MNCP model is the ratio of the assigned link flow to the capacity of each link associated with the existing and predicted maximum flows. The ratio indicates the level of utilization of each individual facility. As shown in Table 3, the links with the value of the ratio close to or more than 1.0 can be identified as saturated links that are approaching or over capacity limit(due to congestion effect). For the test network, links connecting from node 3 to node 5 and from node 5 to node 7 have the ratios with over 0.9 for existing demand. This implies that these links need to be improved in the near future.

<Table 3> Equilibrium Link Flows and Flow/Capacity Ratios Associated with Existing Demand and Predicted Maximum Flow

Mode	Link	x_a	x_a / Q_a	x_a^*	x_a^* / Q_a
Road	1 - 3	70.76	0.89	62.88	0.79
	1 - 4	50.00	0.63	59.52	0.74
	2 - 3	70.00	0.70	76.10	0.76
	2 - 4	50.00	0.63	66.18	0.83
	3 - 5	136.78	1.37	134.38	1.34
	4 - 8	70.48	0.88	78.66	0.98
	5 - 4	20.48	0.34	19.14	0.32
	5 - 6	23.66	0.39	13.53	0.23
	5 - 7	92.64	0.93	101.71	1.02
	6 - 9	73.66	0.92	79.71	1.00
	7 - 8	41.10	0.69	47.57	0.79
	7 -9	47.11	0.79	49.37	0.82
Rail	1 - 3	9.24	0.09	57.60	0.58
	3 - 7	13.22	0.13	62.20	0.62
	7 - 8	8.42	0.08	9.35	0.09
	7 -9	9.24	0.09	57.61	0.58
Intermodal	3	3.98	0.80	4.60	0.92
	7	4.44	0.89	4.77	0.95

Note: x_a= assigned link flow of existing demand, x_a^*= assigned link flow of maximum flow

The information on the current and future levels of capacity utilization of individual facilities is important for transportation agencies and planners since it provides an indication of which links will approach critical congestion, especially when future demand has reached estimated maximum value, and thus need more attention in the planning process.

VI. Conclusions

In this paper, we presented a network capacity model that can be used to assess the capacity of multimodal freight transportation systems at the strategic level of planning. The multimodal network capacity problem was formulated as a nonlinear bi-level optimization problem. Based on a linear approximation technique, a heuristic solution algorithm was developed to solve the MNCP model. Then, the model and its solution algorithm were evaluated for the applicability and effectiveness with a hypothetical multimodal network. The empirical results demonstrated that the MNCP model has been successfully implemented by showing the capability of the model that not only estimate the capacity of multimodal network, but also identify the capacity gaps over all individual facilities in the network, including intermodal facilities. Transportation planners could benefit from the MNCP model in developing sustainable transportation systems in a manner that considers all feasible modes as well as low-cost capacity improvements.

The MNCP model can also be used to examine how, given planned infrastructure investments and other operational changes, adequate and flexible the capacity of a network would be in the future from the multimodal perspective. This could be achieved by using a network-wide sensitivity analysis proposed by Tobin and Friesz (1988) that examines the likely impacts of changes in any of individual system dimensions on network flows. This observation may, in turn, lead us to the development of a capacity-based network design problem that determines the best set of investment options for a multimodal network. In addition, the MNCP model can be extended further to include enhanced realism with respect to modeling real-world networks and freight transportation industry practices. For example, the MNCP model assumes that the flow rate between every origin and destination is fixed and known. In practice, however, these rates may be influenced by the changes in production and consumption activities that are likely to occur in many regions in a long-term time framework. As a result, the assumption of a fix demand pattern should be relaxed to take this variable demand situation into account. In this context, the lower level UE assignment problem embedded in the MNCP model can be replaced by a combined trip distribution, assignment and modal split model such as the one developed by Friesz (1981). Many other possible improvements to the MNCP model remain fruitful topics for future research.

References

1. Cambridge Systematics, Inc. et al. Multimodal Corridor and Capacity Analysis Manual. NCHRP Report 399, TRB, National Research Council, Washington, D.C., 1998.

2. Friesz, T.L. An Equivalent Optimization Problem for Combined Multiclass Distribution, Assignment and Modal Split which Obviates Symmetry Restrictions. Transportation Research Vol. 15B, 1981, pp. 361-369.

3. Ge, Y. E., H. M. Zhang, and W. H. K. Lam. Network Reserve Capacity under Influence of Traveler Information. Journal of Transportation Engineering, May/June 2003, pp.262-270.

4. Guelat, J., M. Florian, and T. G. Crainic. A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows. Transportation Science Vol. 24, 1990, pp. 25-39.

5. Kale, S. R. Intermodal and Multimodal Freight Policy, Planning, and Programming at State Department of Transportation: Review of the Decade Since the Intermodal Surface Transportation Efficiency Act. Transportation Research Record No. 1858, TRB, National Research Council, Washington, D.C., 2003, pp. 69-79.

6. Loureiro, C.F.G and B. Ralston. Investment Selection Model for Multicommodity Multimodal Transportation Networks. Transportation Research Record No. 1522, TRB,National Research Council, Washington, D.C., 1996, pp. 38-46.

7. Morlok, E. K. and S. P. Riddle. Estimating the Capacity of Freight Transportation Systems: A Model and Its Application in Transport Planning and Logistics. Transportation Research Record No. 1653, TRB, National Research Council, Washington, D.C., 1999, pp. 1-8.

8. Park, M. and A. Regan. Capacity Modeling in Transportation: A Multimodal Perspective. Transportation Research Record No. 1906, TRB, National Research Council, Washington, D.C., 2005, pp. 97-104.

9. Tobin, R.L. and T.L. Friesz. Sensitivity Analysis for Equilibrium Network Flows. Transportation Science Vol. 22, 1988, pp. 242-250.

10. Transportation Research Board. Highway Capacity Manual, 3rd Ed.Special Report 209, TRB, National Research Council, Washington, D.C., 1994.

11. Transportation Research Board. Economic Trends and Multimodal Transportation Requirements. NCHRP Report 421, TRB, National Research Council, Washington, D.C., 1999.

12. Transportation Research Board. Highway Capacity Manual, 4th Ed. Special Report 209, TRB, National Research Council, Washington, D.C., 2000.

13. Sheffi, Y. Urban Transportation Network: Equilibrium Analysis with Mathematical Programming Methods. Prentice Hall, Englewood Cliffs, NJ, 1985.

14. Yang, H., M.G.H. Bell, and Q. Meng. Modeling the Capacity and Level of Service of Urban Transportation Networks. Transportation Research, Vol. 34B, 2000, pp. 255-275.