

ANALYSIS OF TRANSPORTATION NETWORK CAPACITY RELATED TO DIFFERENT SYSTEM CAPACITY CONCEPTS

Panatda KASIKITWIWAT
Civil Engineering Division
Faculty of Engineering, Kasetsart University
Kamphaengsaen campus, Nakorn Pathom
73140, Thailand
Email: fengpak@ku.ac.th

Anthony CHEN
Department of Civil and Environmental
Engineering
Utah State University
Logan, UT 84322-4110, USA
Email: achen@cc.usu.edu

Abstract: In transportation, capacity has traditionally been measured at individual elements of the network, such as links (rail lines, road segments and waterway etc.) and nodes (terminals and signalized intersection, etc.). These measures do not constitute the transportation network capacity. Recently, Wong and Yang (1997) proposed the concept of network reserve capacity to estimate the transportation network capacity, but the concept is restricted to a common multiplier for all O-D pairs. In this study, the concepts of ultimate capacity and practical capacity (with a non-uniform O-D growth) are applied to the transportation problem to relax the limitation of the reserve capacity concept; thus, the concepts can yield information regarding the spatial distribution of the demand pattern. The definitions of three different network capacity concepts and their models will be described. The analysis of network capacity for three different concepts is provided to demonstrate their features.

Key Words: Network capacity, bi-level programming, system capacity concepts

1. INTRODUCTION

Network capacity in a transportation system becomes an important measurement for transportation planning and management because it addresses the question of whether or not the transportation system has adequate capacity to handle continuing economic surge and traffic congestion. In transportation, capacity has traditionally been measured at individual elements of the network, such as links (rail lines, road segments and waterway etc.) and nodes (terminals and signalized intersection, etc.). These measures obviously do not constitute the transportation network capacity. However, the maximum network capacity can be determined by the classical maximal flow problem in communication networks, water distribution systems, and electric power systems. But the approach is not directly applicable to a transportation network where capacity modeling characteristics are quite different for the following reasons: (a) the movement in a transportation network involves flows of people rather than pure physical commodities as treated in the classical maximal flow problem, (b) travel delay increases with increasing flow as a result of congestion, as opposed to fixed cost, (c) route choice behavior has to be considered in determining maximum flow of a congested transportation network, (d) the traditional maximal flow problem does not consider level of service when finding the maximum throughput, however, transportation network capacity should be specified with a level of service, such as origin-destination (O-D) travel time, and (e) multiple O-D pairs exist and the flow between different O-D pairs are not exchangeable or substitutable in a transportation network capacity problem. These characteristics make the modeling of a transportation network capacity quite complex, yet intriguing problem to solve (Yang et al., 2000).

Network capacity with a route choice can be expressed mathematically by the bi-level program. The upper-level problem is the network flow maximization and the lower-level problem represents the user travel behavior. The conventional network capacity problems used in various applications, such as optimal signal control or capacity reliability, are based on the concept of network reserve capacity given by Wong and Yang (1997). This concept provides a feasible approach to determine the transportation network capacity with a route choice, but it is restricted to a common multiplier for all O-D pairs. In this study, the concepts of ultimate capacity and practical capacity (with non-uniform O-D growth) are applied to the problem to relax the limitation of the reserve capacity concept; thus they can yield information regarding the spatial distribution of the demand pattern. The problem is also formulated as a bi-level program, in which the upper level maximizes the network capacity subject to roadway and zonal capacity constraints, and the lower level is a combined distribution and assignment model. In this paper, the definitions of three different network capacity concepts and their models are described. Solution algorithms for estimating three different concepts of network capacity are provided. Finally, the analysis of a network capacity for three different concepts is provided to demonstrate their features.

2. NETWORK CAPACITY CONCEPTS

There are two major distinctions at a macro-level of system capacity: economic capacity and physical capacity. Economic capacity is defined as the minimum point on the short-run average cost (AC) curve (i.e., where marginal cost intersects the average cost curve [Scheppach, 1973] or the level of output at which the average total cost [ATC] is minimized [DeLeeuw, 1962]). For physical capacity, various definitions exist that are based on different assumptions about the operation procedures and sustainability of the production rate. In production, the concept of ultimate capacity is the maximum quantity of output which the system can produce, considering only physical limitations on production; the concept of practical capacity is the maximum output at which cost does not exceed a maximum acceptable value (Morlok and Riddle, 1999) or capacity limit, which continues to provide an acceptable level of service deterioration or delay (Kahan, 1979).

In this study, ultimate and practical capacity concepts are applied to the transportation problem to estimate the network capacity for different applications in addition to the existing concept, which is the network reserve capacity concept for estimating network capacity when the zonal growth information is not available. The ultimate capacity concept, when applied to the transportation problem, is the maximum throughput the system can handle without violating roadway and zonal capacity constraints. The network users can then choose both destination and route simultaneously to minimize their user costs. This concept is used to estimate the network capacity of a new developed city. The practical capacity concept, applied to the transportation problem, is the summation of the current O-D demand and the additional demand that the system can accommodate without violating roadway and zonal capacity constraints. This concept is applied to estimate network capacity of an existing city since only additional demands can choose both destination and route simultaneously while the current demand pattern is preserved. For destination choice, additional demands or travelers choose the destination based on travel time to the destination and attractiveness measures of the destination. In system capacity, practical capacity is the maximum output at which cost does not exceed a maximum acceptable value. Therefore, the practical capacity concept in transportation incorporates destination cost for estimating the network capacity.

3. TRANSPORTATION NETWORK CAPACITY MODELS

In this section, models and algorithms for estimating three concepts of network capacity (network reserve capacity, ultimate capacity, and practical capacity) of a transportation network capacity are provided.

3.1 Notation

A	the set of links in the network
N	the set of nodes in the network
I	the set of all origin nodes, $I \subseteq N$
J	the set of all destination nodes, $J \subseteq N$
R	the set of routes in the network
i	an origin node, $i \in I$
j	a destination node, $j \in J$
R_{ij}	the set of routes between origin $i \in I$ and $j \in J$
a	a link in the network, $a \in A$
μ	the O-D matrix multiplier for the whole network
r	a route, $r \in R_{ij}$
C_a	the capacity on link a
v_a	the flow on link a
$t_a(v_a)$	the travel time on link a
\bar{q}_{ij}	the existing demand between O-D pair ij
\tilde{q}_{ij}	the additional demand between O-D pair ij
q_{ij}	the total demand between O-D pair ij , $q_{ij} = \bar{q}_{ij} + \tilde{q}_{ij}$
\mathbf{q}	the O-D demand matrix in vector form
h_r^{ij}	the path flow associated with \bar{q}_{ij}
f_r^{ij}	the path flow associated with \tilde{q}_{ij}
δ_{ar}^{ij}	has a value of 1 if link a is on route r from origin $i \in I$ to destination $j \in J$; 0 otherwise.
\bar{o}_i	the existing trip production at origin i
\tilde{o}_i	the additional trip production at origin i
o_i	the total trip production at origin i , $o_i = \bar{o}_i + \tilde{o}_i$
\bar{d}_j	the existing trip attraction at destination j
\tilde{d}_j	the additional trip attraction at destination j
d_j	the total trip attraction at destination j , $d_j = \bar{d}_j + \tilde{d}_j$
$c_j(d_j)$	cost of destination j
\mathbf{o}	the trip production in vector form
o_i^{\max}	the maximum trip production at origin i (a constant)
d_j^{\max}	the maximum trip attraction at destination j (a constant)
θ	an impedance parameter

3.2 Concept and Model for Network Reserve Capacity

The concept of network reserve capacity with route choice is defined as the largest multiplier μ applied to a given existing O-D demand matrix that can be allocated to a network without violating the link capacities C_a or exceeding a pre-specified level of service (Wong and Yang, 1997). The method for estimating the network capacity uses a common multiplier to scale all O-D pairs. This network capacity model, with a uniform O-D growth, was used in Chen et al. (1999, 2002) to estimate the capacity reliability of a transportation network.

A bi-level program finding the network reserve capacity μ can be formulated as follows:

upper-level problem,

$$\text{Max } \mu \quad (1)$$

subject to,

$$v_a(\mu \mathbf{q}) \leq C_a, \quad \forall a \in A, \quad (2)$$

where $v_a(\mu \mathbf{q})$ is obtained by solving the following route choice problem;

lower-level problem:

$$\text{Min } \sum_{a \in A} \int_0^{v_a} t_a(x) dx \quad (3)$$

subject to,

$$\sum_{r \in R_{ij}} f_r^{ij} = \mu q_{ij}, \quad \forall i \in I, j \in J, \quad (4)$$

$$v_a = \sum_{i \in I} \sum_{j \in J} \sum_{r \in R_{ij}} f_r^{ij} \delta_{ar}^{ij}, \quad \forall a \in A, \quad (5)$$

$$f_r^{ij} \geq 0, \quad \forall i \in I, j \in J, r \in R_{ij}. \quad (6)$$

Route choice behavior and congestion effects are considered by the lower-level problem while the upper-level problem determines the maximum O-D matrix multiplier in Equation (1) subject to the roadway capacity constraints in Equation (2). The lower-level problem is a network equilibrium problem that can be efficiently solved for a given μ . The objective function shown in Equation (3) is the sum of the integrals of the link performance functions. Equation (4) is a set of flow conservation constraints. Equation (5) is the incidence relationships which express the link flows in terms of path flows. Equation (6) represents the non-negativity condition to ensure a meaningful solution. The link flow $v_a(\mu \mathbf{q})$ represents the equilibrium link-flow pattern obtained from solving the lower-level problem for a given existing demand pattern uniformly scaled by μ times the total O-D demands \mathbf{q} . The largest value of μ indicates whether the current network capacity has spare capacity or not. For example, if $\mu > 1$, then the network has a reserve (or spare) capacity amounting to $100(\mu - 1)$ percent of the existing O-D demand matrix \mathbf{q} ; otherwise, the network is overloaded by $100(1 - \mu)$ percent of the existing O-D demand matrix \mathbf{q} .

3.3 Concept and Model for Ultimate Capacity

In general, network capacity is the maximum traffic demand that the system or network can handle. The ultimate capacity concept is defined as the maximum throughput the system can

handle without violating the roadway and zonal capacity constraints. This concept relaxes the common multiplier requirement by allowing the maximum throughput to be scaled by individual O-D pairs. The network capacity model is a variant of the network capacity and the level of service problem described in Yang et al. (2000), which integrates a combined distribution and assignment model to determine the maximum zonal trip productions. This concept allows all travelers in the network to choose both destination and route simultaneously to minimize their cost.

The network capacity model, with a non-uniform growth based on the ultimate capacity concept, is also a bi-level program. The upper level problem maximizes the zonal trip productions subject to the roadway and zonal capacity constraints, while the lower level problem is a combined trip distribution and assignment model. The bi-level program is formulated as follows:

upper-level problem,

$$\text{Max} \sum_{i \in I} o_i \quad (7)$$

subject to,

$$v_a(\mathbf{o}) \leq c_a, \quad \forall a \in A, \quad (8)$$

$$o_i = \sum_{j \in J} q_{ij}(\mathbf{o}) \leq o_i^{\max}, \quad \forall i \in I, \quad (9)$$

$$d_j = \sum_{i \in I} q_{ij}(\mathbf{o}) \leq d_j^{\max}, \quad \forall j \in J, \quad (10)$$

$$o_i \geq 0, \quad \forall i \in I, \quad (11)$$

where $q_{ij}(\mathbf{o})$ and $v_a(\mathbf{o})$ are obtained by solving the combined trip distribution-assignment problem;

lower-level problem,

$$\text{Min} \sum_{a \in A} \int_0^{v_a} t_a(x) dx + \frac{1}{\theta} \sum_{i \in I} \sum_{j \in J} q_{ij} \ln(q_{ij} - 1) \quad (12)$$

subject to,

$$\sum_{j \in J} q_{ij} = o_i, \quad \forall i \in I, \quad (13)$$

$$\sum_{r \in R_{ij}} f_r^{ij} = q_{ij}, \quad \forall i \in I, j \in J, \quad (14)$$

$$v_a = \sum_{i \in I} \sum_{j \in J} \sum_{r \in R_{ij}} f_r^{ij} \delta_{ar}^{ij}, \quad \forall a \in A, \quad (15)$$

$$q_{ij} \geq 0, \quad \forall i \in I, j \in J, \quad (16)$$

$$f_r^{ij} \geq 0, \quad \forall i \in I, j \in J, r \in R_{ij}. \quad (17)$$

Equation (7), in the upper-level problem, determines the maximum total trip productions from all origins subject to the roadway capacity constraints in Equation (8), maximum trip production and attraction constraints in Equations (9) and (10), and non-negativity constraints on trip production in Equation (11). The O-D demand matrix and equilibrium link-flow pattern are solved by the lower-level problem. Both destination and route choices are simultaneously considered in the combined trip distribution-assignment model of the lower-

level problem in Equation (12). Equations (13) and (14) represent the flow conservation constraints. Equation (15) is the incidence relationships that express the link flows in terms of path flows. Equations (16) and (17) are the non-negativity conditions for O-D flows and path flows respectively. The impedance parameter θ for trip distribution in Equation (12) reflects the sensitivity of network users to travel time from an origin to a destination.

3.4 Concept and Model for Practical Capacity

Network capacity with the practical capacity concept is defined as the summation of the current O-D demand and the additional demand that the network can accommodate. This concept, using the network capacity and the level of service problem described in Yang et al. (2000), allows only the additional demand or traveler to choose both route and destination based on the travel cost and the attractiveness measures of destinations, while the current demand pattern is preserved.

In this model, the upper level problem maximizes the additional zonal trip productions subject to the roadway and zonal capacity constraints, while the lower level problem is a combined trip distribution and assignment model with variable destination costs. The bi-level program is formulated as follows:

upper-level problem,

$$\text{Max} \sum_{i \in I} \tilde{o}_i \quad (18)$$

subject to,

$$v_a(\mathbf{o}) \leq c_a, \quad \forall a \in A, \quad (19)$$

$$\tilde{o}_i = \sum_{j \in J} \tilde{q}_{ij}(\mathbf{o}) \leq o_i^{\max} - \bar{o}_i, \quad \forall i \in I, \quad (20)$$

$$\tilde{d}_j = \sum_{i \in I} \tilde{q}_{ij}(\mathbf{o}) \leq d_j^{\max} - \bar{d}_j, \quad \forall j \in J, \quad (21)$$

$$\tilde{o}_i \geq 0, \quad \forall i \in I, \quad (22)$$

where $\tilde{q}_{ij}(\mathbf{o})$ and $v_a(\mathbf{o})$ are obtained by solving the combined trip distribution-assignment problem;

lower-level problem,

$$\text{Min} \sum_{a \in A} \int_0^{v_a} t_a(x) dx + \frac{1}{\theta} \sum_{i \in I} \sum_{j \in J} \tilde{q}_{ij} (\ln \tilde{q}_{ij} - 1) + \sum_{j \in J} \int_0^{\sum_{i \in I} (\tilde{q}_{ij} + \bar{q}_{ij})} c_j(y) dy \quad (23)$$

subject to,

$$\sum_{j \in J} \tilde{q}_{ij} = \tilde{o}_i, \quad \forall i \in I, \quad (24)$$

$$\sum_{r \in R_{ij}} h_r^{ij} = \bar{q}_{ij}, \quad \forall i \in I, j \in J, \quad (25)$$

$$\sum_{r \in R_{ij}} f_r^{ij} = \tilde{q}_{ij}, \quad \forall i \in I, j \in J, \quad (26)$$

$$v_a = \sum_{i \in I} \sum_{j \in J} \sum_{r \in R_{ij}} (f_r^{ij} + h_r^{ij}) \delta_{ar}^{ij}, \quad \forall a \in A, \quad (27)$$

$$\tilde{q}_{ij} \geq 0, \quad \forall i \in I, j \in J, \quad (28)$$

$$f_r^{ij} \geq 0, \quad \forall i \in I, j \in J, r \in R_{ij}, \quad (29)$$

$$h_r^{ij} \geq 0, \quad \forall i \in I, j \in J, r \in R_{ij}. \quad (30)$$

Equation (18), in the upper-level problem, determines the maximum additional productions from all origins subject to the roadway and zonal capacity constraints (Equations (19)-(21)). Equation (22) is the non-negativity constraints for the additional trip productions. The additional O-D demands and route choice are solved in the lower-level problem together with background traffic demands, which are assigned to the network in the conventional Deterministic User Equilibrium (DUE) manner. On the other hand, the additional demand or the traffic growth at each origin zone is distributed among various destination zones by a multinomial logit model, depending on the O-D travel times and the destination costs. The destination cost is an increasing function of the number of trips d_j attracted to the destination j . Equation (23) is the objective function of the Equilibrium Trip Distribution/Assignment with Variable Destination Costs (ETDA-VDC) model, given by Oppenheim (1993). Equations (24)-(26) are the flow conservation constraints for the additional O-D demands, existing path flows, and the additional path flows. Equation (27) is the incidence relationships that express the link flow in terms of path flow. Equations (28)-(30) are the non-negativity conditions for the additional O-D demands, existing path flows, and the additional path flows, respectively. The impedance parameter θ for trip distribution in Equation (23) reflects the sensitivity of network users to travel time from an origin to a destination.

4. SOLUTION ALGORITHMS

In this section, two approaches for solving the three network capacity models are described. The first approach is the incremental assignment-based procedure for solving the network reserve capacity model. The second approach is the genetic algorithm procedure for solving the ultimate and practical network capacity models.

4.1 Solution Algorithm for Solving the Network Reserve Capacity Model

The incremental assignment-based procedure begins by determining an appropriate incremental amount δ of the upper level problem. Then, a standard equilibrium traffic assignment (without capacity constraint) is solved for a given network throughput in order to obtain v_a . These equilibrium link flows are transmitted to the upper-level problem to determine the maximum μ . Because the upper-level problem has only one decision variable, it can be treated as a parameter in the lower-level problem. Then the overall bi-level problem can be solved as a single level problem by varying the value of μ until at least one of the equilibrium link flows violate the capacity constraints (Chen et al., 2002). The lower-level problem uses the linearization algorithm or convex combination method, which was originally suggested by Frank and Wolfe in 1956 as a procedure for solving quadratic programming problems with linear constraints. It is also known as the Frank-Wolfe (FW) method (Sheffi, 1985).

4.2 Solution Algorithm for Solving the Ultimate and Practical Network Capacity Models

The bi-level programming model is intrinsically nonconvex and hence, difficult to solve for a global optimum (Friesz et al., 1990). For the ultimate and practical network capacity models, the fact that link flows and O-D travel time are non-differentiable functions associated with trip production (o_i), causes difficulty in applying the standard optimization approaches for solving the bi-level program. This study adopted the GA method for the ultimate and practical network capacity problems because the GA operates on a population of solutions rather than a single solution, as in most stochastic search methods. In the GA, the best solution is chosen from a number of possible solutions. The approach of the GA implementation for the network capacity problems is that decision variables in the upper level are coded to finite chromosomes and the chromosomes evolve through successive iterations, called generations. During each generation, the chromosomes are evaluated using some measure of fitness, which is calculated by solving the lower level problem. In the GA evolution process, there are three operators including reproduction, crossover, and mutation operators. Reproduction is the selection process of the chromosome from the population set for mating purposes. The crossover operator gets genetic material from the previous generation to the subsequent generation. Mutation is a process that introduces a certain amount of randomness to the search, and enables the search to find solutions that crossover alone may not encounter.

4.2.1 Genetic Algorithm for Solving the Bi-Level Program

- Step 0: Select the sample size and set the number of population (NP), generation (N). Define the crossover and mutation rate.
 - Step 1: Initialization: Code the decision variables in terms of trip production for all origins (\dots, o_i, \dots). Then, set the number of generation $n = 1$.
 - Step 2: Calculate the fitness for individual chromosomes by solving the lower-level problem using Equation (11) for ultimate capacity or Equation (23) for practical capacity
 - Step 3: Carry out the GA evolution process:
 - 3.1: Reproduce the population according to the fitness function values,
 - 3.2: Conduct the crossover operator through a random choice with the designed probability of crossover (p_c),
 - 3.3: Conduct the mutation operator through a random choice with the designed probability of mutation (p_m).
- This step yields a new population at generation $n + 1$
- Step 4: If $n =$ the maximum number of generations, the sample with the highest fitness is adopted as an approximated optimal solution of the problem. Otherwise, set $n = n + 1$ and go to Step 2.

To handle the roadway capacity constraint in the upper-level problem, the fitness function should incorporate the constraint violations by means of penalty method. The fitness function with the penalty factor ρ is described as follows:

$$F(x) = f(x) - \rho \left[\max \left(0, \sum_{a \in A} (v_a - C_a) \right) \right].$$

4.2.2 Algorithm for Solving the Lower-level Problem of the Ultimate Capacity Model

The solution algorithm is based on the partial linearization method, which is a descent algorithm for continuous optimization problems (Patriksson, 1994). Only a part of the objective is linearized in each iteration. The procedure of solving the combined trip distribution and assignment model is as follows:

Step 0: Find a set of feasible flow $\{v_a^k\}, \{q_{ij}^k\}$, and set $k = 1$.

Step 1: Calculate link cost $t_a^k = t_a(v_a^k), \forall a$.

Step 2: Find the direction:

2.1: Calculate the minimum travel-time path from each origin i to all destinations, based on $\{t_a^k\}$. Let c_{ij}^k denote the minimum travel time from origin i to destination j ,

2.2: Determine the auxiliary O-D flow by applying a logit distribution model, that is,

$$d_{ij}^k = \frac{o_i e^{-\theta c_{ij}^k}}{\sum_m e^{-\theta c_{im}^k}}, \forall i \in I, j \in J,$$

2.3: Assign d_{ij}^k to the minimum travel-time path between origin i and destination j .

This also yields a link-flow pattern $\{y_a^k\}$.

Step 3: Determine the step size. Find α_k that minimizes the function

$$\text{Min}_{0 \leq \alpha_k \leq 1} \sum_{a \in A} \int_0^{v_a^k + \alpha_k (y_a^k - v_a^k)} t_a(x) dx + \frac{1}{\theta} \sum_{i \in I} \sum_{j \in J} [q_{ij}^k + \alpha_k (d_{ij}^k - q_{ij}^k)] [\ln(q_{ij}^k + \alpha_k (d_{ij}^k - q_{ij}^k)) - 1]$$

Step 4: Update O-D and link flows. Set

$$q_{ij}^{k+1} = q_{ij}^k + \alpha_k (d_{ij}^k - q_{ij}^k), \forall i \in I, j \in J,$$

$$v_a^{k+1} = v_a^k + \alpha_k (y_a^k - v_a^k), \forall a \in A,$$

Step 5: Test the convergence. If convergence is not achieved, set $k = k+1$ and go to step 1.

Otherwise, terminate: the solution is $\{v_a^{k+1}\}, \{q_{ij}^{k+1}\}$.

4.2.3 Algorithm for Solving the Lower-level Problem of the Practical Capacity Model

The partial linearized algorithm given by Yang et al. (2000) can be summarized below:

Step 0: Determine an initial value $\{v_a^k\}, \{\tilde{q}_{ij}^k\}$ and set $k = 0$.

Step 1: Calculate link cost $t_a^k = t_a(v_a^k)$, the minimum travel time from origin i to destination j , $c_r^{ij(k)}$ and destination cost $c_j^k = c_j(\sum_i (\tilde{q}_{ij}^k + \bar{q}_{ij}))$.

Step 2: Find the descent direction by obtaining $\tilde{a}_r^{ij(k)}$ that minimizes sub problem P1, which is

$$\text{Min } Z_1(x) = \sum_{i \in I} \sum_{j \in J} \sum_{r \in R_{ij}} \left[c_r^{ij(k)} + \frac{1}{\theta} \ln(\tilde{q}_{ij}^k) + c_j (\sum_{i \in I} \tilde{q}_{ij}^k + \bar{q}_{ij}) \right] \tilde{a}_r^{ij(k)}$$

subject to,

$$\sum_{j \in J} \sum_{r \in R_{ij}} \tilde{a}_r^{ij} = \tilde{o}_i, \forall i \in I, \tilde{a}_r^{ij} \geq 0, \forall i \in I, j \in J, r \in R_{ij},$$

and $\bar{a}_r^{ij(k)}$ that minimizes program P2, which is

$$\text{Min } Z_2(y) = \sum_{i \in I} \sum_{j \in J} \sum_{r \in R_{ij}} c_r^{ij(k)} \bar{a}_r^{ij(k)}$$

subject to,

$$\sum_{r \in R_{ij}} \bar{a}_r^{ij} = \bar{q}_{ij}, \forall i \in I, j \in J, \bar{a}_r^{ij} \geq 0, \forall i \in I, j \in J, r \in R_{ij}.$$

$$\text{Then, set } y_a^k = \sum_{i \in I} \sum_{j \in J} \sum_{r \in R_{ij}} (\tilde{a}_r^{ij(k)} + \bar{a}_r^{ij(k)}) \delta_{ar}, \forall a \in A, \tilde{d}_{ij}^k = \sum_{r \in R_{ij}} \tilde{a}_r^{ij(k)}, \forall i \in I, j \in J.$$

Step3: Determine the step size. Find α_k that minimizes the function

$$\begin{aligned} \text{Min } Z(\alpha) = & \sum_{0 \leq \alpha_k \leq 1} \int_0^{v_a^k + \alpha_k (y_a^k - v_a^k)} t_a(x) dx + \frac{1}{\theta} \sum_{i \in I} \sum_{j \in J} [\tilde{q}_{ij}^k + \alpha_k (\tilde{q}_{ij}^k - \tilde{d}_{ij}^k)] \\ & \times [\ln(\tilde{q}_{ij}^k + \alpha_k (\tilde{q}_{ij}^k - \tilde{d}_{ij}^k)) - 1] + \sum_{j \in J} \int_0^{\sum_{i \in I} \tilde{q}_{ij}^k + \bar{q}_{ij} + \alpha_k (\tilde{q}_{ij}^k - \tilde{d}_{ij}^k)} c_j(y) dy. \end{aligned}$$

Step4: Update O-D and link flows. Set

$$\tilde{q}_{ij}^{k+1} = \tilde{q}_{ij}^k + \alpha_k (\tilde{d}_{ij}^k - \tilde{q}_{ij}^k), \forall i \in I, j \in J,$$

$$v_a^{k+1} = v_a^k + \alpha_k (y_a^k - v_a^k), \forall a \in A.$$

Step 5: Test the convergence test. The condition of convergence is

$$\left| \frac{\tilde{q}_{ij}^k}{o_i} - \frac{\exp\{-\theta[c_{ij}^k + c_j (\sum_{i \in I} \tilde{q}_{ij}^k)]\}}{\sum_{m \in J} \exp\{-\theta[c_{im}^k + c_m (\sum_{i \in I} \tilde{q}_{im}^k)]\}} \right| \leq \varepsilon, \forall i \in I.$$

If convergence is not achieved, set $k = k+1$ and go to step 1. Otherwise, terminate and using the solution $\{v_a^{k+1}\}, \{\tilde{q}_{ij}^{k+1}\}$.

5. NUMERICAL EXAMPLE

A simple network given in Figure 1 consists of six nodes, seven links, two origins, two destinations and four O-D pairs. Existing O-D demands for O-D 1-3, O-D 1-4, O-D 2-3 and O-D 2-4 are 40, 10, 10, and 50 respectively. The link travel time function used is the standard Bureau of Public Road (BPR) function:

$$t_a = t_a^f \left(1 + 0.15 \left(\frac{v_a}{C_a} \right)^4 \right),$$

where v_a , t_a^f , and C_a are the flow, free-flow travel time, and capacity on link a , respectively. The values of free-flow travel time and link capacities are shown in Table 1.

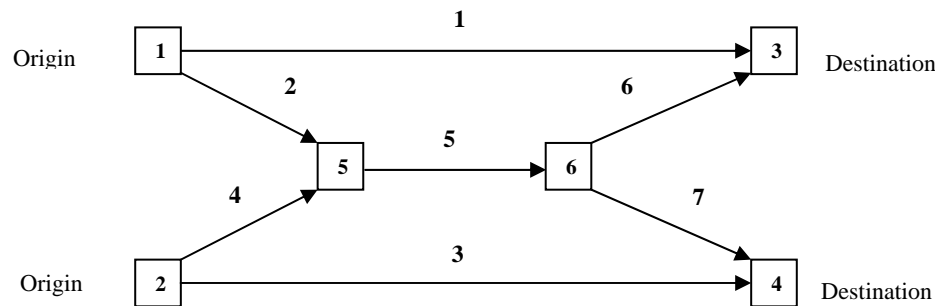


Figure 1 Test Network

Table 1 Link Free-Flow Travel Times and Link Capacities

Link #	Free-flow travel time	Link capacity
1	10.00	100.00
2	4.00	80.00
3	12.00	80.00
4	4.00	50.00
5	5.00	120.00
6	5.00	50.00
7	4.00	50.00

5.1 Network Reserve Capacity for Different Demand Pattern

The results presented in Table 2 provide the maximum network reserve capacity that the system can accommodate. In the network reserve capacity concept, traffic demand in every O-D pair increases at the same rate, and different patterns of O-D demand (with the same total demand) can be increased differently according to link capacity constraints. The pattern where O-D demands are more congruous with the network topology would achieve a higher network capacity. Among three different patterns experimented with, pattern 1, which utilizes link 1 (with the highest initial flow on O-D (1-3)) more than other patterns, gives the highest network reserve capacity, which is 227.92 with the increased rate of 2.072.

Table 2 Existing O-D Demand Patterns and Their Network Reserve Capacity

O-D	Initial Pattern 1	O-D flow Result	Initial Pattern 2	O-D flow Result	Initial Pattern 3	O-D flow Result
1-3	40	82.88	30	61.20	25	41.65
1-4	10	20.72	20	40.80	25	41.65
2-3	10	20.72	20	40.80	30	49.98
2-4	50	103.60	40	81.60	30	49.98
Network capacity	110.00	227.92	110.00	224.40	110.00	183.26

5.2 Ultimate Capacity for Different Impedance Parameter Values

The ultimate capacity concept allows all users in the network to have both destination and route choices. The impedance parameter θ for the trip distribution reflects the sensitivity of network users to travel time from an origin to a destination. An increase in impedance parameter θ would generate an O-D travel demand/or trip length frequency with a shorter O-D travel time. Figure 2 shows the results of the ultimate network capacity model and the O-D flow patterns for different impedance parameters. For this specific network, values of the impedance parameter from 0.01 to 6 are tested. With a low impedance parameter value, the network shows the uniform allocation of flows among the O-D pairs. With a higher impedance parameter value, flows on O-D (1-3) and O-D (2-4), which have lower O-D travel times, are higher. However, flow on O-D (1-3) stops increasing due to the roadway capacity constraint of link 1 when the impedance parameter value reaches 0.5. The pattern that best fits with the network topology will give the highest network capacity. For this network, the pattern with an impedance parameter value of 0.25 gives the highest network capacity.

5.3 Practical Capacity for Different Existing Traffic Demands

The network capacity estimation, based on the practical capacity concept, considers the cost in addition to the limitations of link and zonal capacities. Road users base their decisions of making a trip on travel cost and destination cost. In this problem, destination cost is incorporated in the model and therefore only the additional demands or travelers have the destination and route choices. Thus, the pattern of the existing demand has to be preserved; different levels of existing demand would affect the amount of additional demands that can be added to the network without violating the roadway and zonal capacity constraints. Table 3 provides the parameters for the destination cost function. Table 4 presents the results of the network capacity for three different levels of existing traffic demand. Pattern 1 from the network reserve capacity concept is the base level. The other two levels are 25% lower and 25% higher than the base level. Among the three levels of existing demand, the highest network capacity can be obtained from the case of the lowest existing demand. Because the existing demand pattern has to be preserved, the higher level of existing demands means that lower additional demands could have both destination and route choices; thus, the network capacity with a higher level of existing demand is lower than the network capacity with a lower level of existing demand.

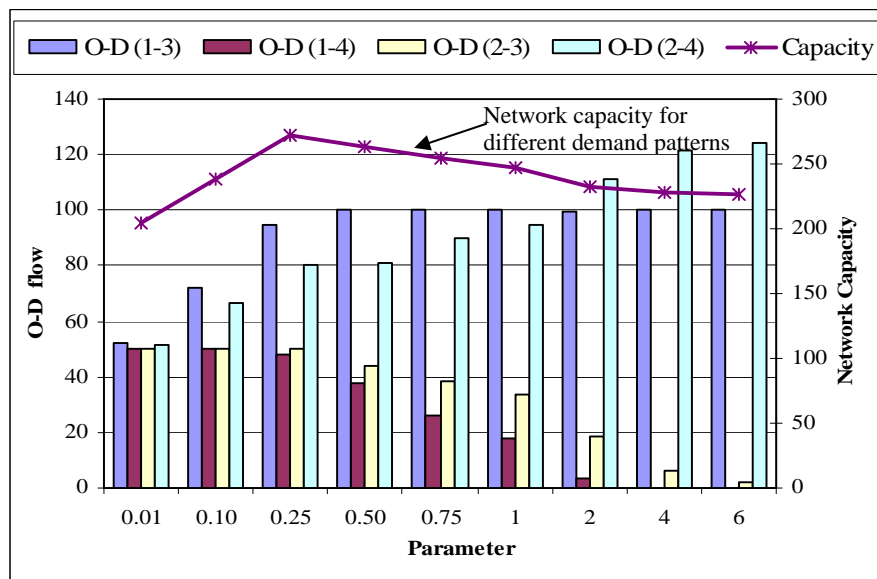


Figure 2 Network capacity and O-D Distributions for the Ultimate Capacity Concept

Table 3 Destination Cost Data for Test Network $c_j(d_j) = \alpha_j d_j^{\beta_j} - m_j$ for Practical Model

Destination	m_j	α_j	β_j
3	1.20	0.15	0.25
4	1.50	0.10	0.25

Table 4 Network Capacities for Different Levels of Existing Traffic Demand

Practical Capacity Concept	System Capacity	Trip Production		Origin-Destination Flow			
		O ₁	O ₂	O-D(1-3)	O-D(1-4)	O-D(2-3)	O-D(2-4)
Existing demand (0.75 * base case)	-	-	-	30.00	7.50	7.50	37.50
Flow results	260.72	138.34	122.38	100.82	37.52	34.54	87.84
Existing demand (base case)	-	-	-	40.00	10.00	10.00	50.00
Flow results	257.58	137.83	119.75	100.80	37.03	31.36	88.39
Existing demand (1.25 * base case)	-	-	-	50.00	12.50	12.50	62.50
Flow results	254.87	134.50	120.37	99.79	34.71	29.28	91.09

5.4 Comparison of Network Capacity for Different Concepts

This section compares the network capacity results using three different network capacity models. For the network reserve capacity concept, a pre-determined O-D pattern is required. This pattern remains fixed while the network capacity model determines the maximum O-D demand multiplier that can be accommodated when assigned to the network using a user-equilibrium method without exceeding the roadway capacity constraints. For the ultimate capacity concept, O-D flows do not need to follow a pre-determined pattern since a combined distribution and assignment model is adopted in the bi-level program to determine both

destination and route choices of all users. However, physical zonal capacities (o_i^{\max}) are required in the upper level problem.

Network capacity under this model is treated as an ultimate capacity, which is the upper bound because all network users can choose both destination and route simultaneously. For the practical capacity concept, an existing O-D demand pattern is preserved. Only the additional demands can choose both destination and route. To choose the destination, network users also consider cost of destination in addition to the cost of traveling to the destination. Therefore, a combined distribution and assignment model with destination cost is applied only to the additional demands. The complete results of these three models are provided in Table 5 and graphically presented in Figure 3, to highlight the distinct feature of the three network capacity concepts. For the network capacity model using the reserve capacity concept, the maximum network capacity is 227.92 (or the scalar multiplier is equal to 2.072). The pre-determined O-D pattern is preserved with a uniform growth rate of 2.072 for all O-D pairs. For the ultimate network capacity model (an impedance parameter value of 0.5) that allows both destination and route choices for all network users, the maximum network capacity is 262.54, an increase of 15.19 % compared to the network reserve capacity model. For the practical capacity concept that allows destination choice for the additional demands, the network capacity model with destination cost and an impedance parameter value of 0.5 gives the maximum network capacity of 257.58, an increase of 13.08 % compared to the network reserve capacity model. It is 1.89 % less than the network capacity of the ultimate capacity concept.

Moreover, it is observed that the O-D patterns resulting from the three capacity models are quite different. For example, the results indicate that the flows on O-D (1-3), O-D (1-4), and O-D (2-3) from the practical and ultimate capacity models, are higher than those in the reserve capacity model, while flows on O-D (2-4) show the opposite. However, the net increase in network capacity is larger in the network capacity models that allow partial and full destination choice. Network capacity estimated with the reserve capacity concept is underestimated because the same increase rate is applied to all O-D pairs. If flows on one O-D pair cause some links in the network to reach the roadway capacities, flows on other O-D pairs will also stop increasing. In this network, link 3, which served O-D (2-4), is the bottleneck link. The practical and ultimate capacity models can achieve more network capacity by having more demands in O-D (1-3) (Link 1), O-D (1-4) (Links 2-5-7), and O-D (2-3) (Links 4-5-6) since the capacities of these routes serving these O-D pairs are underutilized in the reserve capacity model. Note that the volume-to-capacity (V/C) ratio of link 4, for the practical capacity, is less than the V/C ratio for the reserve capacity because link 4 is also used by flows on O-D (2-4) in the reserve capacity model.

Table 5 Network Capacities and O-D Demands for Three Different Concepts

Reserve capacity concept (Uniform O-D growth)	System Capacity	Trip Production		Origin-Destination Flow			
		O ₁	O ₂	OD(1-3)	OD(1-4)	OD(2-3)	OD(2-4)
Input	-	-	-	40.00	10.00	10.00	50.00
Flow results (multiplier = 2.072)	227.92	103.60	124.32	82.88	20.72	20.72	103.60
Ultimate capacity concept (Non-uniform O-D growth)	System Capacity	Trip Production		Origin-Destination Flow			
		O ₁ ^{max}	O ₂ ^{max}	OD(1-3)	OD(1-4)	OD(2-3)	OD(2-4)
Inputs	-	150.00	150.00	-	-	-	-
Flow results	262.54	138.01	124.53	99.97	38.04	43.75	80.78
Practical capacity concept (Non-uniform O-D growth)	System Capacity	Trip Production		Origin-Destination Flow			
		O ₁ ^{max}	O ₂ ^{max}	OD(1-3)	OD(1-4)	OD(2-3)	OD(2-4)
Input (existing traffic pattern)	-	150.00	150.00	40.00	10.00	10.00	50.00
Flow results	257.58	137.83	119.75	100.80	37.03	31.36	88.39

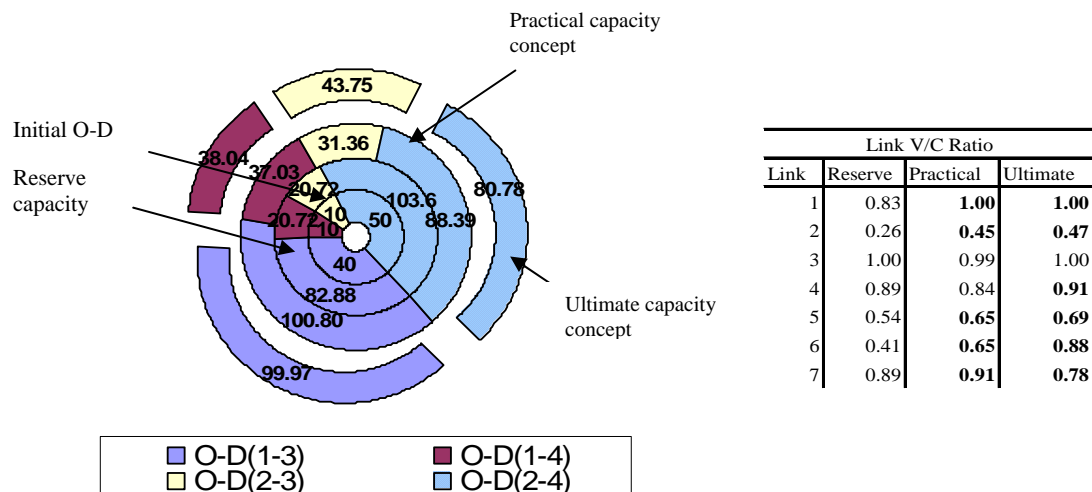


Figure 3 Comparison of Network Capacity Concepts

6. CONCLUSION

In this paper, three network capacity models and solution approaches for solving these network capacity models are provided. Numerical results are provided to illustrate how each network capacity model estimate the capacity of the network. For the first model, network capacity determination is based on the premise of preserving a pre-determined O-D pattern and the capacity is determined by scaling all O-D pairs using a common multiplier. This restriction is relaxed by adopting two non-uniform network capacity models for two different concepts: the ultimate capacity concept and the practical capacity concept. This allows a non-uniform O-D growth in the spatial distribution of the O-D demand pattern. It also corrects the under-estimate biased problem in the network capacity model by allowing a partial and full destination choice. Different concepts of network capacity models provided in this study can be applied to the transportation problem for different applications. Network reserve capacity is for estimating network capacity when the zonal growth information is not available. The ultimate capacity concept is used to estimate the network capacity of the new developed city and the practical capacity is applied to estimate the network capacity of an existing city.

ACKNOWLEDGMENT

This research was supported by the NSF CAREER Grant: CMS-0134161.

REFERENCES

- Bell, M.G.H. and Iida, Y. (1997) **Transportation Network Analysis**. John Wiley & Sons, New York.
- Chen, A., Yang, H., Lo, H.K. and Tang, W. (1999) A capacity related reliability for transportation networks. **Journal of Advanced Transportation Vol. 33, No. 2**, 183-200.
- Chen, A., Yang, H., Lo, H.K. and Tang, W. (2002) Capacity reliability of a road network: an assessment methodology and numerical result. **Transportation Research B Vol. 36, No.3**, 225-252.
- DeLeeuw, F.(1962) The concept of capacity. **The American Statistical Association Journal**, December, 826-839.
- D'Este, G. M, and Taylor, M A P. (2003) Network vulnerability: an approach to reliability analysis at the level of national strategic transport networks. In Bell, M G H and Iida, Y. **The Network Reliability of Transport**. Elsevier-Pergamon, Oxford.
- Friesz, T.L., Cho, H.J., Mehta, N.J., Tobin, R.L. and Anandalingam, G. (1992) A simulated annealing approach to the network design problem with variational inequality constraints. **Transportation Science Vol. 26**, 18-26.
- Kahan, A. M. (1979) Railway capacity analysis and related methodology. Canadian Transport Commission Research Branch.
- Morlok, E. K., and Riddle, S.M. (1999) Estimating the capacity of freight transportation systems: a model and its application in transport planning and logistics. **Transportation Research Record Vol. 1653**, 1-8.
- Oppenheim, N. (1993) Equilibrium trip distribution/assignment with variable destination costs. **Transportation Research B Vol. 27 No.3**, 207-217.
- Patriksson, M. (1994) **The traffic assignment problem: models and methods**. VSP, Utrecht, The Netherlands.
- Scheppach, R C.1973. Transportation Capacity Volumes 1-5. Jack Faucett Associates Inc.
- Sheffi, Y. (1985) **Urban transportation networks: equilibrium analysis with mathematical programming methods**. Prentice-Hall, Englewood Cliffs, New Jersey.
- Wong, S.C., and Yang, H. (1997) Reserve capacity of a signal-controlled road network. **Transportation Research B Vol. 31 No. 5**, 397-402.
- Yang, H., Bell, M.G.H. and Meng, Q. (2000) Modeling the capacity and level of service of urban transportation networks. **Transportation Research B Vol. 34 No. 4**, 255-275.