Transit Network Design with Variable Demand

Young-Jae Lee[1] and Vukan R. Vuchic[2]

Abstract: This paper shows the iterative approach to solving transit network design problem, particularly with variable transit demand under a given fixed total demand. Although recent studies, which use a simplified combinatorial search approach, showed their capability of building optimal transit networks and handling the complicated transit travel time characteristics, only this iterative approach is believed to properly handle the dynamic characteristics of the relationship between variable transit trip demand and optimal transit network design. Since transit demand depends on the configuration of the transit network and frequencies of the routes, this approach is more desirable for transit network planning than combinatorial approach. The basic approach generates the optimal transit network from the initial network, which requires the shortest in-vehicle travel time, through iterating the assignment procedure and the improvement procedure until there is no more improvement in the network. With variable transit demand, the modal split procedure is added to the basic model to generate the optimal transit network and to estimate transit demand simultaneously. This paper also shows the relationship between optimal transit network design and critical design inputs, such as transit operating speed, total demand size, and transfer penalty. As results of the analysis, synergistic effect of variable transit demand and the optimal transit network are discussed.

DOI: 10.1061/(ASCE)0733-947X(2005)131:1(1)

CE Database subject headings: Transportation networks; Design; Travel demand.

Introduction

In order to provide better service to users and to increase operating efficiency, transit system planning should produce transit services that provide competitive travel time and require low operating costs. Because of complex transit travel time characteristics, which include in-vehicle travel time, waiting time, transfer time, and transfer penalties, it has been a difficult task to optimize transit networks.

Numerous scholars, including Newell (1979) and Baaj and Mahmassani (1991), have pointed out that traditional mathematical programming has difficulties in generating an optimal transit network due to the reasons including nonlinearity and nonconvexity of the model, combinatorial explosion, multiobjective nature, and spatial layout of routes.

Recently, with improvement of search algorithms and computer technology, important heuristic research has been done (Hasselström 1981; Baaj and Mahmassani 1991; Ceder and Israeli 1998; Pattnaik et al. 1998; Shi et al. 1998; Chien et al. 2001). All of those studies are based on the combinatorial search approach.

One key point of the combinatorial approach is efficient generation of sample spaces, which are candidate routes and candidate sets of routes. Depending on the generated sample spaces, the optimality of the results is basically decided, even if an improvement procedure follows. Also, the number of generated candidate routes and that of candidate sets of routes are critical in this method. If those numbers are too big, then this method becomes close to the all-enumeration method. If they are too small, it is hard to generate good routes and sets of routes for the sample spaces. Thus, this approach tends to rely on the network designer's knowledge to obtain a good simplified sample space. Consistency and generalization of the network designer's knowledge are required as well.

The other key point is the flexibility of the methodology in respect to handling constraints. Although the combinatorial search approach may be able to give good results with given fixed inputs, it is not flexible enough to include certain dynamic inputs, particularly those such as variable transit demand.

Only Rea (Rea 1971) used the iterative approach for transit network design, which uses transit travel time characteristics. Among transit travel time components, in-vehicle travel time and waiting time have a tradeoff relationship. If a transit network provides a direct connection, it gives shorter in-vehicle travel time to users, but may require longer waiting time due to the reduced amount of demand per route. On the other hand, if a transit network consists of circuitous routes and/or requires transfers, it may require longer in-vehicle travel time and/or transfer time, but will provide shorter waiting time due to the higher frequencies of routes resulted from the concentrated demand per route. Although Rea's study contains important concepts, there are some difficulties in using this method in the real world. First, this study uses individual links instead of an integrated set of links as a single route. So, this methodology can only be considered as the transit version of highway assignment. Second, even though service frequency is changed with the amount of demand, a limited number of predefined frequency sets are used for a wide range of demand.

The present study shows the iterative approach to solving the transit network design problem. This approach is flexible enough

[1]Assistant Professor, Institute for Transportation, Morgan State Univ., 5200 Perring Pkwy., Baltimore, MD 21251. E-mail: yjlee@eng.morgan.edu

[2]UPS Foundation Professor of Transportation, Dept. of Electrical and Systems Engineering, Towne Building, Univ. of Pennsylvania, Philadelphia, PA 19104-6315.

Note. Discussion open until June 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on September 10, 2002; approved on April 13, 2004. This paper is part of the ***Journal of Transportation Engineering***, Vol. 131, No. 1, January 1, 2005. ©ASCE, ISSN 0733-947X/2005/1-1–10/$25.00.

to deal with dynamic characteristics of transit network design. In particular, this paper focuses on how to deal with variable transit demands under a given fixed total demand. To execute this methodology, the computer software, transit network designer (*TRANED*), was programmed with C++. The relationship between produced optimal transit networks and input elements will also be discussed.

Basic Model

The study starts from the "basic model" for transit network design, which includes minimum constraints to show the algorithm effectively. This basic model can be expanded with various realistic constraints, and in the later section, the comprehensive model, including a constraint of the variable transit demand, will be shown.

Objective

The objective of the algorithm is to build an optimal transit network. Three general objectives in defining an optimal network are as follows:

- user travel time minimization, measured in person hours;
- transit agency's profit maximization, measured in dollars; and
- social benefit maximization or social cost minimization.

User travel time minimization is usually the objective of public transit ownership. However, a transit agency's cost and revenue should be considered as constraints. For a private transit agency, profit maximization is the main objective of transit network design, but as a constraint, user travel time should be considered. Social benefit maximization or social cost minimization is a combination of the above two objectives, and it is the common objective under public transit ownership. In this case, while two objectives are used as the combined objective, there are no such constraints used under the previous two objectives. In recent years, multiobjective algorithms have been used for transit network design (Baaj and Mahmassani 1991; Cedar and Israeli 1998). In this case, multiple solutions are generated, which have different user's travel costs and operator's costs. Among those solutions, the best combination of those two objectives is chosen by a planner.

In this paper, user travel time minimization is used as the optimization criterion for simplicity. However, as previously mentioned, this model can include various agency's operational constraints such as fare box recovery ratio. If the operator's constraints are satisfied, user travel time minimization is desirable in many situations of public ownership.

Scheduling Process

In order to estimate passenger waiting time, it is necessary to define the service frequency of each route. Generally, frequency is determined by the supply ability of the transit agency, passenger demand, and/or headway policy. Supply frequency is limited by the fleet size and/or fare box recovery ratio. Policy headway usually sets the minimum frequency of a route. In most real world cases, all three frequencies are considered in determining actual frequency. Although all those frequencies can be considered and dealt with as constraints, for simplicity, demand frequency is used to determine actual frequency in this paper.

The "demand frequency" is estimated based on the volume of users. This frequency is considered the minimum frequency that provides just enough capacity to satisfy the demand on the maximum load section (MLS) so that demand is always less than capacity on the other sections (links) of the route. The demand frequency f_D is shown in the following equation (Vuchic et al. 1976; Cedar and of Israeli 1998):

$$f_D = \frac{V_{\mathrm{MLS}}}{C_v \cdot \alpha}, \tag{1}$$

where f_D=frequency decided by demand (vehicle/h); V_{MLS}=volume on the maximum load section (person/h); C_v=vehicle capacity (space/vehicle); and α=load factor (person/space).

Fig. 1 shows how the demand in the MLS is determined. Demand on each section is computed as difference of accumulated boarding passengers and alighting passengers. If it is assumed that vehicle capacity is 50 spaces and load factor is 0.85, then the demand frequency f_D is 18.8 vehicles/h with the computed V_{MLS} of 800 persons/h. This frequency may be rounded up to 20 vehicles/h for scheduling simplicity and convenience.

Algorithm

Unlike auto travel, which increases auto travel time with increased auto travel demand due to congestion, increased transit travel demand decreases transit travel time due to the higher service frequency. However, in order to have more transit riders under fixed transit demand, circuitous routing is unavoidable. It results from a tradeoff relationship between in-vehicle travel time and waiting time in a transit network. The methodology of this paper starts from this "concentration of flow" concept, which was introduced and used by Rea (1971) and Hasselström (1981), although they limited its usage at the realization and applications as mentioned.

The iterative approach in this paper looks for the minimum total travel time network starting from generating the minimum in-vehicle travel time network. Then the transit network is gradually improved by increasing in-vehicle travel time while decreasing waiting time. This algorithm consists of three major steps: generation of an initial network, assignment, and network improvement. They are followed by a supporting step, network analysis. These steps are iterated until the optimal transit network is generated as shown in Fig. 2. The generated optimal transit network provides direct connections to major travel flows, while also providing shorter waiting times to minor travel flows by generating circuitous travel paths.

The first step involves generating the initial network with the minimum number of routes using the shortest path algorithm (Dijkstra 1959; Whiting and Hillier 1960; Dantzig 1966). This provides minimum in-vehicle travel time paths to all origin–destination pairs. For this procedure, the shortest paths for all origin–destination pairs are generated; included paths are then eliminated to avoid unnecessary overlapping paths.

The second step repeats the transit assignment procedure, which concentrates transit travel flow to certain routes. This procedure allows higher frequencies of certain routes and shorter total travel time. As a result, less efficient routes are eliminated from the network.

The third step improves the transit network through changing the alignments of routes. After building an initial network and adjusting it to assignment procedure, some alignment changes of certain routes for the improvement of the network should be considered for reducing users' travel times. After stabilizing frequencies of routes in the transit network through repeated assignment

(trip)

Station	1	2	3	4	5	6	7	8	9	10
Boarding	50	100	250	250	300	200	300	200	100	0
Alighting	0	0	50	100	150	100	250	250	350	500
Accumulated boarding	50	150	400	650	950	1150	1450	1650	1750	1750
Accumulated alighting	0	0	50	150	300	400	650	900	1250	1750
Demand on each section	50	150	350	500	650	750	800	750	500	0

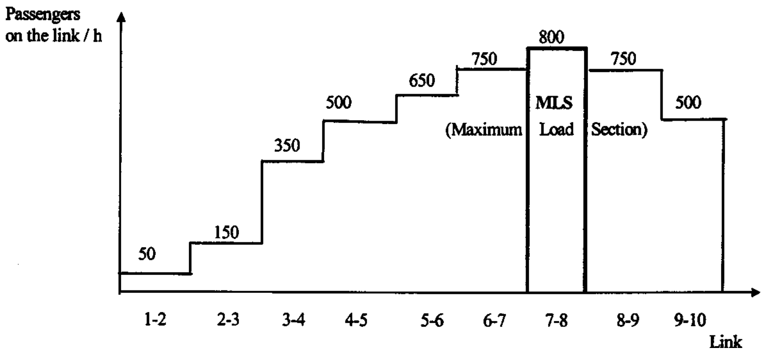


Fig. 1. Passenger assignment along route

procedures, routes are reviewed and alignments are changed where necessary. Since less frequent routes require longer waiting times that cause longer travel times, they would be considered first. Since the network consists of selected routes, routes in Baaj and Mahmassani's initial network may need to be split and changing branches in addition to merging routes (Baaj and Mahmassani 1991). However, the procedure in this analysis merges routes and removes unused nodes for network improvements, because the initial network of this study starts from all shortest travel time routes.

For merging routes, there are two cases: one is merging routes which have shared trucks and same-directed branches; the other is merging routes which have shared trucks and opposite-directed branches. If branches of two routes are going from the same station of the shared trunk section, then it is called same-directed branches. If branches of two routes are going from different stations of the shared trunk section, then it is called opposite-directed branches.

Network analysis is the supporting step to generate outputs resulting from the above steps. The outputs of each step, such as number of routes, total travel time, and frequency of routes, are compared to those of the previous step.

The results of this procedure were generated and compared with other researches (Baaj and Mahmassani 1991; Mandle 1979), in order to prove the validation of the methodology (Lee 1998). The results show that transit networks generated by *TRANED* generally require less travel time for users.

This basic model is simple; however, because of the flexibility of the mathematical programming of the iterative approach, this methodology can add various realistic constraints to the basic model. Additional constraints to those in the basic model are operational and financial constraints, coordination with existing service (intermodal coordination), express service, schedule information for users, and variable transit demand.

Inputs and Outputs for Model

To provide a model for different cities and conditions, it is therefore necessary to develop a general model for a transit network. Required input elements for the model are as follows:

- template network (basic network with links and nodes);
- origin–destination travel demand;
- distance or in-vehicle travel time on each link by mode;
- transit unit (TU) capacity of given mode;
- relative weight for waiting time compared to in-vehicle travel time;
- transfer penalty; and
- relative weight for transfer time compared to in-vehicle travel time.

For the purpose of analyzing the network generated by *TRANED*, the following network characteristics are also computed by *TRANED* in addition to the basic output - network configuration and frequencies of routes.

- network configuration or route configurations (-);
- frequencies of routes [vehicle/h];
- total in-vehicle travel time in the network (person min/h);
- total waiting time in the network (person min/h);
- total transfer time in the network (person min/h);
- total transfer penalties in the network (person min/h);
- total travel time in the network (person min/h);
- total travel time except in-vehicle travel time (person min/h);
- travel demand without transfer (persons);
- travel demand requiring transfer (persons);
- total travel demand (persons);
- degree of circuity (%);
- number of routes (-);
- total route length in the network (km);
- average route length (km); and
- total vehicle operational time in the network (vehicles min/h).

Most of the outputs are self-explanatory, but some require additional explanation. The degree of circuity is the parameter showing the indirectness of travel. There are two types of circuities: physical and time. While physical circuity represents circuity of routes, time circuity represents circuity of travel. The main differences between the two are transfer time and penalty. While physical circuity does not include transfer time and penalty as extra costs, time circuity considers them as extra costs due to the indirectness of a route. Time circuity is used in this study. Time circuity is the ratio of the extra travel time after boarding a transit vehicle due to the indirectness of routes, possible transfer time, and transfer penalties to the shortest in-vehicle travel time as the

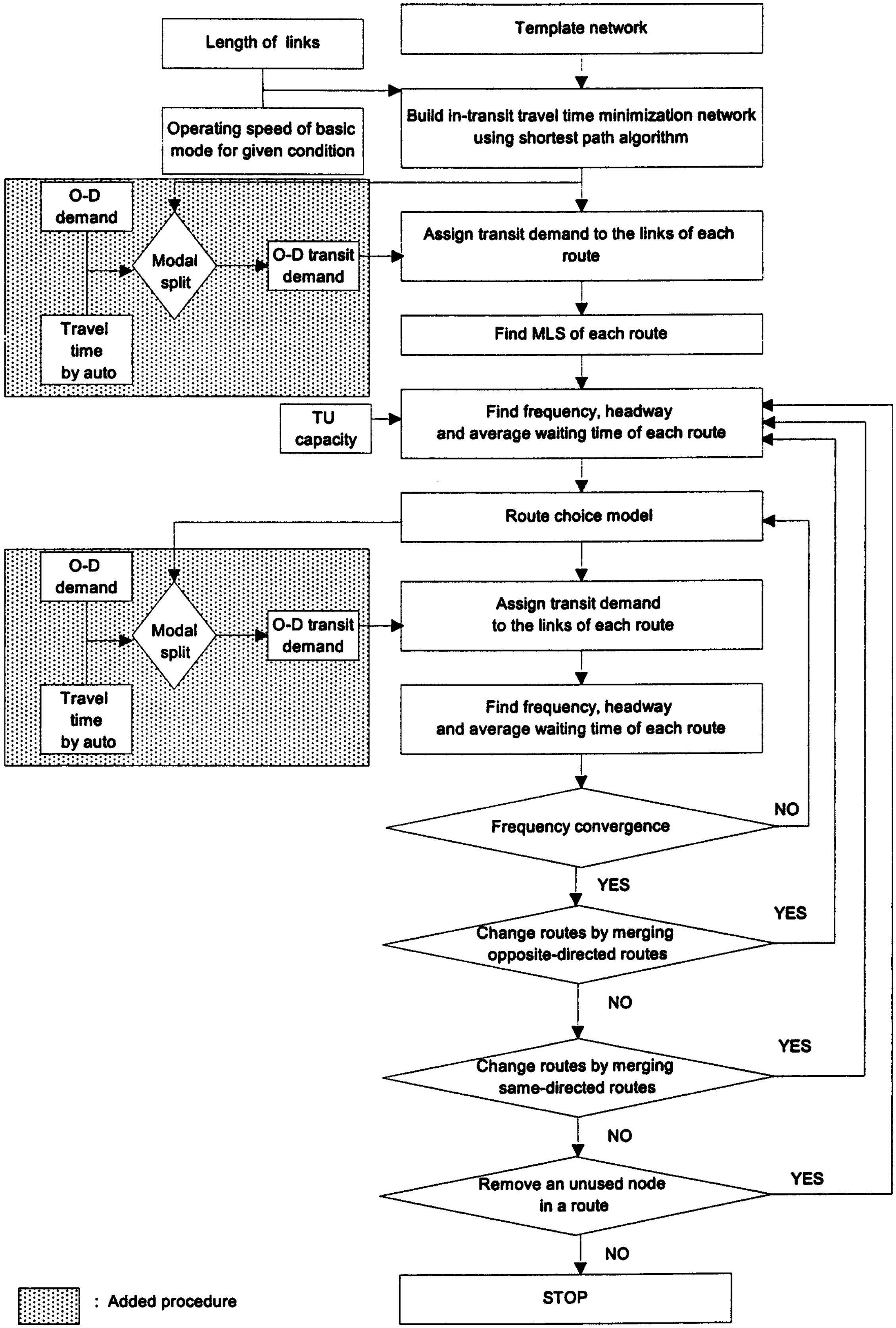


Fig. 2. Final procedure for transit network design for basic model

following equation. Degree of circuity (DOC) in the network is the average of an individual user's degree of circuity

$$\mathrm{DOC}(\%) = 100 \cdot \frac{\Delta t_t + t_i + p}{\min t_i} \tag{2}$$

where Δt_i=additional in-vehicle travel time (difference between real in-vehicle travel time and in-vehicle travel time of shortest path); t_t=transfer time; p=transfer penalty; and min t_i=in-vehicle travel time of shortest path.

Total vehicle operational time (TOT) in the network is the total vehicle operating time in the network. It is calculated as follows. The "2" in the equation means two-directional service, which is conventional in most transit service

$$\mathrm{TOT} = 2\sum_k f_k l_k \tag{3}$$

where k=route number; f=frequency; and l=length of route as minutes or operating time for one direction.

Variable Transit Demand with Fixed Total Demand

While the basic model uses fixed given transit demand, application with variable transit demand will be shown in this section. Variable demand consists of two kinds of variations. One is variable transit demand due to changes in the modal split between auto and transit under given total demand, and the other is variable total demand, which may result from the feedback process of the urban transportation planning process (UTPP) (FHWA/UMTA 1977). Although variable total demand can be applied to the model, for simplicity, application of variable transit demand is the focus of this paper.

Variable transit demand is defined by the following data; origin–destination total travel demand, in-vehicle travel time by different modes (auto and transit), and a rule for the modal split to determine transit demand from the total demand.

Total demand can be estimated from the first and second steps of the UTPP, which are trip generation and trip distribution. However, since both are generated based on a given highway and transit network, estimation of total demand is not realistic when the transit network does not yet exist. To solve this problem, transit network design should be a part of the UTPP, so that generating total demand, transit demand, and transit network should be developed together. This total package of planning requires a complicated procedure.

Although travel time of each origin–destination pair in the basic model was variable and depended on the network configuration and frequencies of routes, in-vehicle travel times on the links of a transit mode were given as fixed due to the fixed transit demand. However, those in-vehicle travel times could also be variable in this realistic model as a function of travel volume.

Travel times and demands of participating modes are all related to feedback processes as follows. Basically, transit in-vehicle travel time is dependent upon auto in-vehicle travel time, and auto in-vehicle travel time is dependent upon auto demand volume. Auto demand volume depends on the ratio between auto travel time and transit travel time, which depend on their in-vehicle travel times. Transit in-vehicle travel time depends on transit network and auto in-vehicle travel time.

Since this process includes auto assignment to the network, it is extremely complex. For simplicity, it is assumed that auto in-vehicle travel times on the links and transit in-vehicle travel times on the links are given, although they can be easily converted to a variable in the algorithm. Even with these simplicities, transit travel time and transit demand are still variable depending on the transit network design.

In addition to total demand and travel time of each mode, a rule for the modal split is necessary. With this rule and travel time of each mode, transit demand and auto demand can be estimated from the given total demand. For estimating the modal split, the logit model is the most popular by far. The logit formulation is a share model that divides the persons between the various modes depending on each mode's relative desirability for any given trip (Khisty and Lall 1998, pp. 494–497). The probability of using mode i, P_i, is given by

$$P_i = \frac{e^{U(i)}}{\sum_{r=1}^{n} e^{U(r)}}, \tag{4}$$

where $U(i)$=utility of mode i; $U(r)$=utility of mode r; and n=number of modes in consideration.

Utility of each mode, U_i, can be calibrated by the following equation. Since inputs for the equation are disutilities (costs), U_i has a negative value. Calibrated values of coefficients depend on the conditions of the applied area of the model

$$U_i = -a_i - b_i X - c_i Y - d_i C \tag{5}$$

where U_i=utility function of mode i; a_i, b_i, c_i, d_i=coefficients of mode i; X=in-vehicle travel time; Y=out-of-vehicle time; and C=cost of travel.

Fig. 3 shows the entire procedure of the revised model, which adds the procedure of determining transit demand to the basic model shown in Fig. 2.

Example

In order to generalize the example, inputs of Rea's paper are basically applied to this paper. The network used in this example has 16 nodes and they are connected to each other as shown in Fig. 4(a). Link travel times and origin–destination total travel demand for this example are modified from Rea's inputs, because they are too short and small to make a reasonable example. The doubled travel times of the links estimated from the distances and operating speed (about 30 km/h) of Rea's example are also shown on the template network. Fig. 4(b) shows total trip demand, and it is ten times the amount of origin–destination travel demand of Rea's example.

As other input elements for the model, TU capacity, transfer penalty, and relative weight for waiting time and transfer time must be defined. For TU capacity, 60 spaces are assumed as in Rea's paper. For simplicity, no transfer penalty will be applied. That means there are no additional fares, additional access time, and other qualitative inconveniences related to transfers, but transfer waiting is still applied. As a relative weight of the waiting time to in-vehicle travel time, the ratio of 1 will be used, which means the values of waiting time and in-vehicle travel time are the same.

As a modal split rule, a simplified logit model is used. Auto in-vehicle travel time is assumed as 0.8 times transit in-vehicle travel time and 0.05 is assumed the coefficient b_{auto}, and also the same number is assumed for other coefficients (b_{transit} and c_{transit}) of both in-vehicle travel time and out-of-vehicle travel time for transit mode. No out-of-vehicle travel time of auto is assumed ($Y_{\text{auto}}=0$) and that of transit includes waiting time and transfer

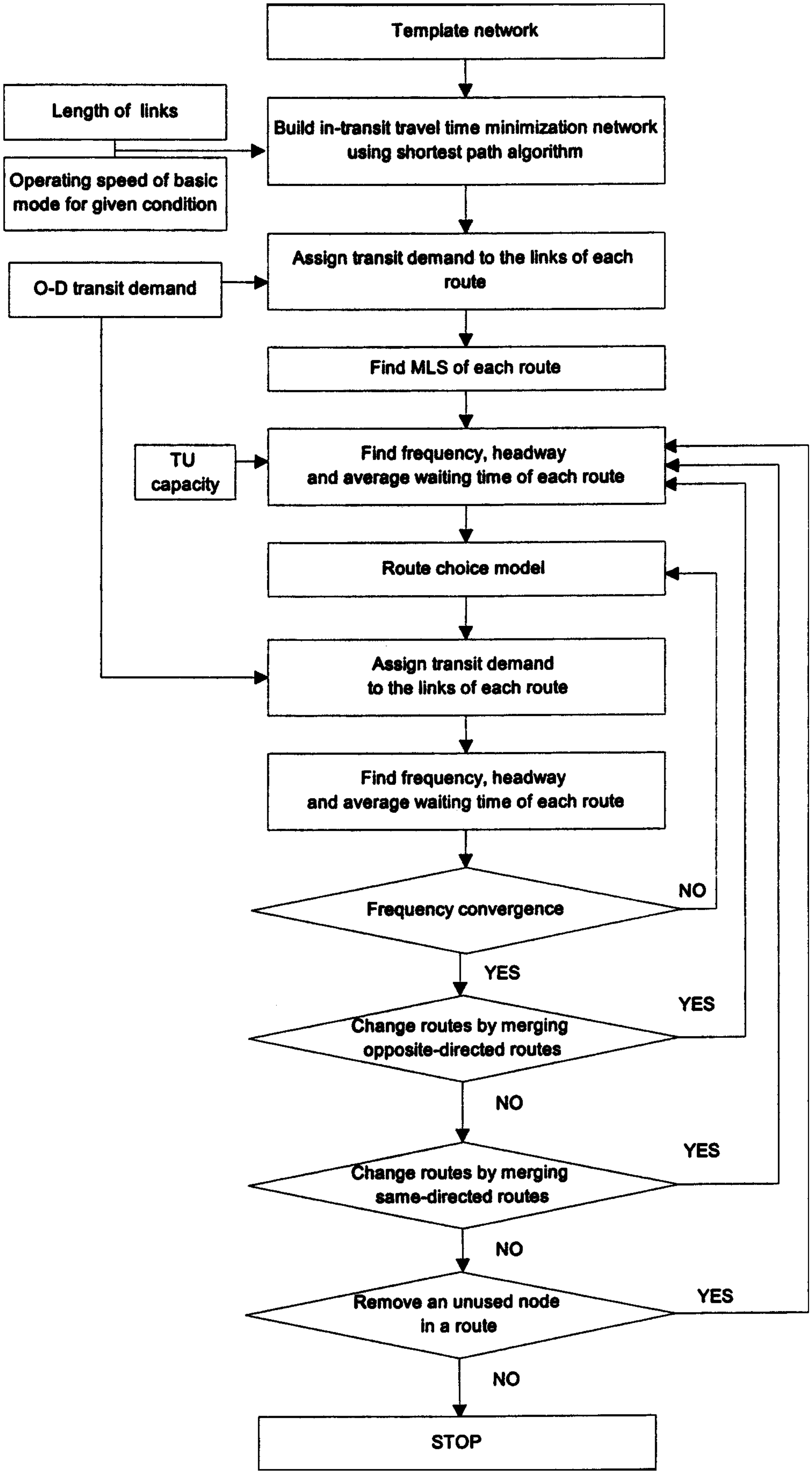
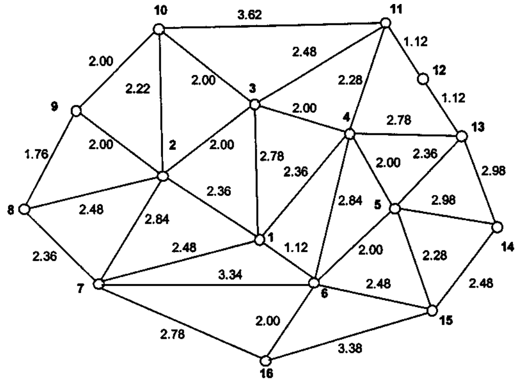


Fig. 3. Procedure for transit network design with variable transit demand



(a) Template network of the Rea's example and in-vehicle travel time (minutes)

(trip/hour)

Node	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10	# 11	# 12	# 13	# 14	# 15	# 16
# 1	0	50	50	50	50	50	0	0	0	0	0	0	0	0	0	0
# 2	300	0	200	300	200	200	100	100	100	100	100	100	100	100	100	100
# 3	300	200	0	300	200	200	100	100	100	100	100	100	100	100	100	100
# 4	50	50	50	0	50	50	0	0	0	0	0	0	0	0	0	0
# 5	300	200	200	300	0	200	100	100	100	100	100	100	100	100	100	100
# 6	300	200	200	300	200	0	100	100	100	100	100	100	100	100	100	100
# 7	400	100	100	400	100	100	0	50	50	50	50	50	50	50	50	50
# 8	400	100	100	400	100	100	50	0	50	50	50	50	50	50	50	50
# 9	400	100	100	400	100	100	50	50	0	50	50	50	50	50	50	50
# 10	400	100	100	400	100	100	50	50	50	0	50	50	50	50	50	50
# 11	400	100	100	400	100	100	50	50	50	50	0	50	50	50	50	50
# 12	400	100	100	400	100	100	50	50	50	50	50	0	50	50	50	50
# 13	400	100	100	400	100	100	50	50	50	50	50	50	0	50	50	50
# 14	400	100	100	400	100	100	50	50	50	50	50	50	50	0	50	50
# 15	400	100	100	400	100	100	50	50	50	50	50	50	50	50	0	50
# 16	400	100	100	400	100	100	50	50	50	50	50	50	50	50	50	0

(b) Origin-Destination total demand

Fig. 4. Example for study

time. The other inputs in the equation, which are a_{auto}, a_{transit}, C_{auto}, and C_{transit}, are assumed to be zero for simplicity.

Results of Example

For the initial network, a total of 28 routes are generated, as shown in Table 1. It took 31 iterations to optimize the example and estimate transit demand from the given total demand. Table 2 shows the routes of selected iterations, which "survive" after each iteration. As discussed before, the assignment procedure only removes routes and does not change the alignment of routes. However, the procedure of merging routes does change route alignments. Results of travel time components and other factors of selected iterations are shown in Table 3 and the final network configuration and route frequencies are shown in Table 4. To optimize this example, only the assignment procedure and the procedure for merging routes were used. The procedure for the treatment of unused nodes was not necessary for this specific case.

As shown in Table 3, resulting from *TRANED*, average total travel time (per trip) consisting of three travel time components is improved from 12.69 to 7.18 min. In the meantime, in-vehicle travel time increases while other components decrease. This indirectness is also shown at the "degree of circuity," which increases through the iterations to collect travel flows and to increase frequencies for certain trips.

Transit demand starts from 12,734 trips initially based on the shortest in-vehicle travel time and no waiting time. Although transit demand is dropped to 10,901 trips with increased in-vehicle travel time and waiting time at the second iteration, transit demand starts to increase with improved transit network. As a result of the reduced transit travel time, final transit demand (TD) is increased from 10,901 trips of the first iteration to 11,792 trips

Table 1. Initial Routes of Example

Route number	Nodes
1	5-6-1-2-9
2	3-1-6-16
3	7-1-4-11-12
4	1-7-8
5	10-2-1-6-15
6	7-1-4-13
7	9-2-1-6-15-14
8	10-2-1-6-16
9	4-3-2-8
10	3-2-7
11	8-2-3-11-12-13
12	3-4-5-14
13	4-3-2-9
14	5-4-3-10
15	3-4-5-15
16	11-4-6-16
17	5-6-7-8
18	11-4-5-15
19	12-13-5-6-16
20	9-8-7-16
21	7-2-10
22	8-7-6-15-14
23	1-2-8
24	8-9-10
25	9-10-11-12-13
26	10-11-12-13-14
27	12-13-5-15
28	14-15-16

among 25,800 trips of total demand, which is 45.7% of the total demand. During the network improvement procedure, the number of routes in the network is reduced from 28 to 5, and total route length in the network is shortened from 195.1 to 45.5 min. For the better presentation of changes in major outputs throughout the iterations, Fig. 5 is provided. As another indicator, the transit additional travel time ratio (TATTR) is shown. This indicator represents the additional travel time when transit is used compared to auto travel time. It is calculated with the following equation:

$$\text{TATTR}(\%) = \frac{\dfrac{\sum_i \sum_j T_{ij}^t}{\sum_i \sum_j D_{ij}^t} - \dfrac{\sum_i \sum_j T_{ij}^a}{\sum_i \sum_j D_{ij}^a}}{\dfrac{\sum_i \sum_j T_{ij}^a}{\sum_i \sum_j D_{ij}^a}} \cdot 100 \tag{6}$$

where i,j=nodes; t=transit; a=auto; T=total travel time; and d=demand.

Since auto in-vehicle travel time is assumed to be 0.8 times the transit in-vehicle travel time, if there is no waiting time and transfer time for transit trips, and direct routes are provided as the initial network, TATTR is 25%. In Fig. 5, TATTR starts from 263.6% at the initial network due to the long waiting time, but goes down to 106.0% at the final network. In the meantime, the number of routes (NOR) in the network settles at five.

Table 2. Results of Iterations: Changes in Routes of Example

Iteration number	Procedure	Route numbers
Initial		1–28
1	Assignment	1, 2, 3, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 19, 23
2	Assignment	1, 2, 3, 6, 7, 8, 11, 12, 13, 14, 16
3	Assignment	1, 2, 3, 6, 7, 8, 11, 12, 14, 16
4	Assignment	1, 3, 7, 6, 8, 11, 12, 14, 16
5	Assignment	1, 3, 7, 6, 8, 11, 12, 14
		↓
8	Assignment	1, 3, 6, 7, 8, 11, 12
		↓
14	Merging	3, 6, 7(9-2-1-6-5-15-14), 8, 11, 12
		↓
26	Merging	3(7-1-4-11-12-13), 7(9-2-1-6-5-15-14), 8, 11, 12
		↓
31	Assignment	3(7-1-4-11-12-13), 7(9-2-1-6-5-15-14), 8, 11, 12

Sensitivity Analysis

In this section, three major inputs are examined to show their relationship to the generated optimum transit networks: (1) difference between auto travel time and transit travel time, which are the main factors to decide utilities of modes; (2) total demand size; and (3) transfer penalty.

Before undertaking a sensitivity analysis of a transit network, several difficulties in analysis should be pointed out. First, since the network configurations and their frequencies are generated by heuristics, they do not represent an exact optimum and the results provided by a generated network may be inconsistent in terms of closeness to the optimum. This heuristic method may weaken the relationship between inputs and outputs. Second, the number of routes and the lengths of routes are discrete, and that causes an inconsistent and discontinuous relationship between inputs and outputs. Third, there are inputs used in the model, and the sensitivity depends on what the values of those inputs are. If a dominant input, which influences the results most, such as demand volume and distribution, is given in different ranges, then the sensitivity analysis may not be reliable or consistent. For these reasons, the results of sensitivity analysis may not provide exact statistical values, but they will show the trends of the relationship between inputs and outputs of the network. If those trends are reasonable, then this algorithm can be considered to react soundly with various inputs.

Changes in Transit Speed

While auto travel time is assumed to be 0.8 times the transit in-vehicle travel time in Fig. 5, different ratios of auto travel time to transit in-vehicle travel time, with values of 0.4, 0.6, 1.0, and 1.2, are applied. Estimated transit demand shares with different ratios are plotted in Fig. 6(a). As predicted, with increased transit operating speed (decreased in-vehicle travel time), TD increases from 43.7% with 0.4 travel time ratio to 47.8% with 1.2 travel time ratio. Consequently, the TATTR decreases from 303.6% (0.4 travel time ratio) to 39.7% (1.2 travel time ratio), and the NOR in the network increases from 5 to 6.

When transit operating speed increases, TATTR decreases dra-

Table 3. Results of Iterations: Network Characteristics of Example

Iteration number	Average in-vehicle travel time (min/trip)	Average waiting time (min/trip)	Average transfer time (min/trip)	Average total travel time (min/trip)	Transit demand (trips)	Number of routes	Degree of circuity	Total route length (min)	Total vehicle operating time (vehicles min/h)
Initial	4.36	8.33	0.00	12.69	12,734	28	0.00	195.1	1,648.3
1[a]	4.58	4.89	0.49	9.96	10,901	17	16.28	123.6	1,641.2
2[a]	4.75	2.92	0.68	8.35	11,245	11	24.54	82.7	1,780.9
3[a]	4.79	2.70	0.72	8.21	11,513	10	26.38	76.7	1,776.6
4[a]	4.81	2.41	0.73	7.95	11,552	9	27.06	70.8	1,767.3
5[a]	4.83	2.11	0.66	7.60	11,567	8	25.92	63.7	1,833.4
				↓					
8[a]	4.93	1.88	0.67	7.48	11,691	7	28.44	57.7	1,880.0
				↓					
14[b]	5.17	1.64	0.63	7.44	11,700	6	33.03	52.0	2,004.6
				↓					
26[b]	5.10	1.50	0.59	7.19	11,738	5	30.50	45.5	1,928.3
				↓					
31[a]	5.09	1.50	0.59	7.18	11,792	5	30.28	45.5	1,922.1

[a]Improvement by assignment.
[b]Improvement by merging routes.

Table 4. Results of Iterations: Final Network Configuration and Frequencies

Route number	Configuration	Frequency (vehicles/h)
3	7-1-4-11-12-13	27.5
7	9-2-1-6-5-15-14	29.4
8	10-2-1-6-16	12.2
11	8-2-3-11-12-13	14.8
12	3-4-5-14	16.2

matically due to a synergistic effect. This is because of decreased in-vehicle travel time of transit, as well as decreased waiting time and transfer time due to increased demand with reduced total travel time by transit.

Changes in Total Demand Size

Fig. 6(b) shows the changes in the transit network characteristics with different total demand from 50 to 200% of the total demand of the basic case. With increased total demand, the absolute amount of transit demand increases. Due to the increased transit demand and resulting higher frequencies, transit networks become more efficient, and this efficiency increases TD share. So with increased total demand, not only does the amount of transit demand increase, but also TD share increases due to a synergistic effect (43.6, 45.7, 46.5, and 46.9%, respectively). Consequently, the NOR increases (4, 5, 6, and 7, respectively) due to greater transit demand and greater TD share. Also, TATTR decreases (162.1, 106.0, 84.4, and 75.9%, respectively) for the same reasons.

Changes in Transfer Penalty

In Fig. 6(c), different amounts of transfer penalties are applied to the *TRANED* with variable transit demand. With increased transfer penalties (from 0 to 10, 20, and 30 equivalent minutes of in-vehicle travel time), TD share decreases from 45.7% to 29.9, 18.4, and 12.3%. Consequently, TATTR, which includes transfer penalties as an equivalent time of in-vehicle travel time, increases from 106.0% to 188.1, 272.0, and 398.3%, respectively. Increased transit additional travel time is affected not only by transfer penalties, but also by the decreased TD share and the network configurations generated less efficiently to avoid transfer penalties.

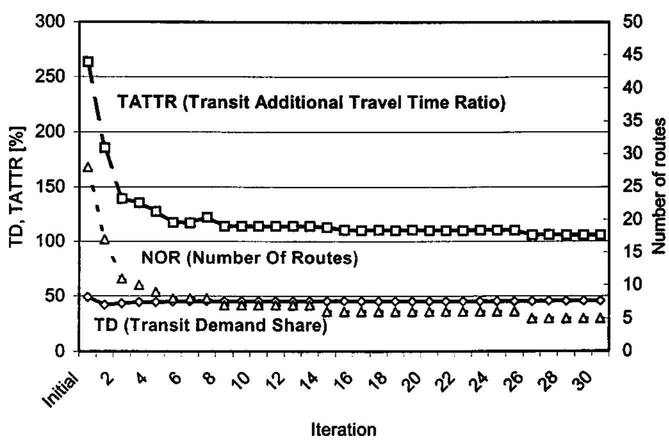


Fig. 5. Transit network characteristics with variable transit demand

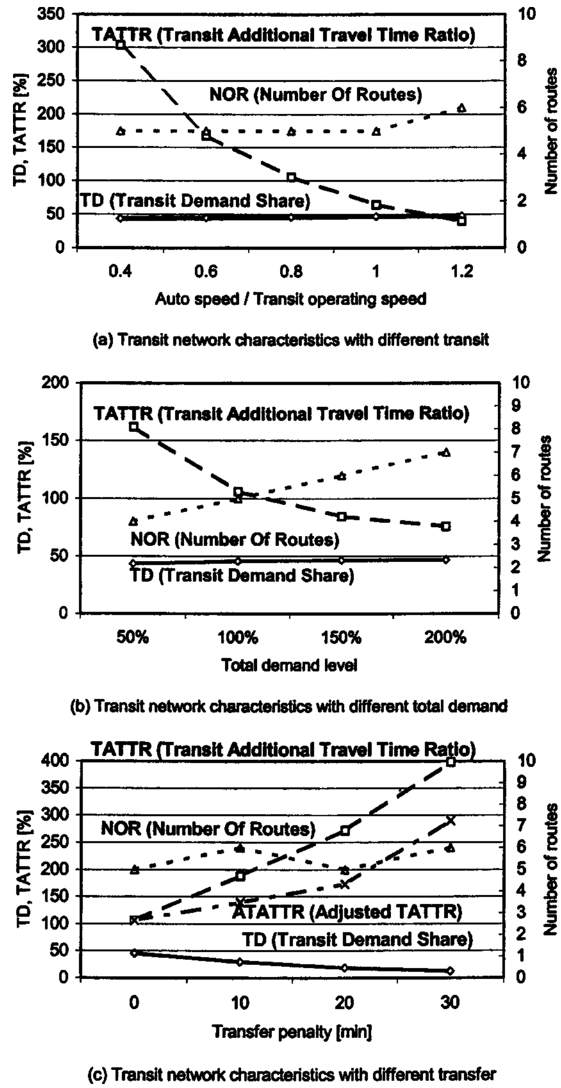


Fig. 6. Results of sensitivity analysis

Fig. 6(c) also shows that adjusted transit additional travel time ratio (ATATTR), which does not include transfer penalties, also increases (from 106.0% to 139.3, 173.0, and 290.1%) due to reduced TD share and the transit network generated less efficiently, although it is not as dramatic as that of TATTR.

The (NOR) in the network shows inconsistency for the reason explained previously in this section—inconsistency due to using heuristics for the sensitivity analysis. With increased transfer penalties, the configuration of transit network is changed in two different ways: fewer and more circuitous routes with less demand, and more direct and higher number of routes with more demand (Shih et al. 1998). Depending on the other inputs, a certain output can be inconsistent. When transfer penalty is increased from zero to 10 min, the NOR increased from 5 to 6, because of sufficient demand to make a direct service. However, when transfer penalty increased from 10 to 20 min, the number of routes decreases from 6 to 5 to make concentrated flows to increase frequencies of routes due to less demand. In the case of transfer penalty of 30 equivalent min of in-vehicle travel time, the number of routes should not be increased to make an efficient transit network because of decreased demand with increased transfer penalties.

As a result of the increased number of routes in the network, which is the inconsistent result of *TRANED*, the increase of both the TATTR and the adjusted ratio (ATATTR) are much higher with 30 min of transfer penalty compared to lesser transfer penalties.

Conclusion

In this paper, the application of the iterative transit network design for the variable transit demand was examined. Since this application for the variable transit demand has distinct dynamic characteristics of the transit network design, this iterative approach can handle the procedure more efficiently than other combinatorial approaches.

The relationships with generated networks and the changes in input elements, such as transit operating speed, total travel demand, and transfer penalty, were also examined for sensitivity analysis. With those changes, not only the network characteristics such as configuration and frequencies, but also transit demand share could be estimated by the iterative method.

It is well known that higher transit demand can generate more efficient transit networks due to the resulting higher frequencies of routes in the network. In this paper, furthermore, it was shown that there are synergistic effects between variable transit demand and generated optimal transit network. If an input is changed in favor of transit with faster transit operating speed or reduced transfer penalty, as a result, not only by those favorable inputs, but also by the increased transit demand resulting from those changes, the transit network becomes much more efficient with a higher number of routes and less total travel time for transit users.

References

Baaj, M. H., and Mahmassani, H. S. (1991). "An AI-based approach for transit route system planning and design." *J. Adv. Transp.*, 25(2), 187–210.

Cedar, A., and Israeli, Y. (1998). "User and operator perspective in transit network design." *Proc. 77th Annual Meeting of TRB*, Paper No. 980267, Washington, D.C., January.

Chien, S., Yang, Z., and Hou, E. (2001). "Genetic algorithm approach for transit route planning and design." *J. Transp. Eng.*, 127(3), 200–207.

Dantzig, G. B. (1966). "All Shortest Routes in a Graph." *Proc., Int. Symp. Théorie des Graphes*, Rome, Dunod, Paris, 91–92.

Dijkstra, E. W. (1959). "A note on two problems in connection with graphs." *Numer. Math.*, 1, 269–271.

Federal Highway Administration and Urban Mass Transportation Administration (FHWA/UMTA). (1977). *An introduction to urban travel demand forecasting: A self-instructional text*, United States Department of Transportation, Washington, D.C.

Hasselström, D. (1981). "Public transportation planning." PhD dissertation, Univ. of Gothenburg, Gothenburg, Sweden.

Khisty, C. J., and Lall, B. K. (1998). *Transportation engineering: An introduction*, 2nd Ed., Prentice-Hall, Upper Saddle River, N.J.

Lee, Y.-J. (1998). "Analysis and optimization of transit network design with integrated routing and scheduling." PhD dissertation, Univ. of Pennsylvania, Philadelphia.

Mandle, C. E. (1979). "Evaluation and optimization of urban transportation networks." *Proc., presented at the 3rd European Congress on Operations Research*, Amsterdam, The Netherlands.

Newell, G. F. (1979). "Some issues relating to the optimal design of bus routes." *Transp. Sci.*, 13(1), 20–35.

Pattnaik, S. B., Mohan, S., and Tom, V. M. (1998). "Urban bus transit network design using genetic algorithm." *J. Transp. Eng.*, 124(4), 368–375.

Rea, J. C. (1971). "Designing urban transit systems: An approach to the route-technology selection problem." *No. UMTA-URT-49(70)-71-6*, Urban Transportation Program, Univ. of Washington, Seattle.

Shih, M., Mahmassani, H. S., and Baaj, M. H. (1998). "A planning and design model for transit route networks with coordinated operations." *Proc., 77th Annual Meeting of TRB*, Paper No. 980418, Washington, D.C.

Vuchic, V. R., et al. (1976). *Transit operating manual*, Pennsylvania Department of Transportation, Univ. of Pennsylvania, University Park, Pa.

Whiting, P. D., and Hillier, J. A. (1960). "A method for finding the shortest route through a road network." *Oper. Res. Q.*, 11, 37–40.