Issues Related to Transit Network Design Problem

Mahmoud Owais
Assistant Professor at Civil
Faculty of Engineering, Civil
Department Assiut

ABSTRACT
Public Transit (P.T) is very important means to reduce traffic congestions, to improve urban environmental conditions and consequently affects people social lives. Planning, designing and management of P.T are the key issues for offering a competitive mode that can compete with the private transportation. These transportation planning, designing and management issues are addressed in the Transit Network Design Problem (TNDP). It deals with a complete hierarchy of decision making process. It includes strategic, tactical and operational decisions. The main body of TNDP is two stages, namely; route design stage and frequency setting. The TNDP is extensively studied in the last five decades; however the research gate is still widely open due to its many practical and modeling challenges. In this paper, a comprehensive background is given to illustrate the issues and challenges related to the TNDFP to help in directing the incoming researches towards the untouched areas of the problem.

General Terms
Transportation – Computer applications

Keywords

1. INTRODUCTION
The transit planning and operation process commonly includes five basic activities, usually performed in sequence: (1) network route design, (2) frequency setting, (3) timetable development, (4) vehicle scheduling, and (5) crew scheduling [1-3]. The output of each activity positioned higher in the sequence becomes an important input for lower-level decisions. In practical, first and second component are combined under the Transit Network Design Problem (TNDP).

TNDP is the most important component in Transit planning, in which the overall cost of the public transportation system highly depends on it. The TNDP aims to design a set of bus routes and manage these routes operation in an efficient manner for both users and operators. Different system functions and targets, required for each group of participants, have to be met through the solution methodology. TNDP, stated simply, relates to the determination of a set of routes defined over the street network with their corresponding schedules to deal with demand trips [4].

There are some basic concepts should be clarified before proceeding on. These concepts are recognized by transit research community. They would help in getting a good grasp of the problem.

NOMENCLATUR

d_{i,j} \quad \text{the transit demand from (i) to (j) expressed as trips per unit time}
t_{i\to j} \quad \text{total user travel time between (i) and (j) = (waiting + in-vehicle time)}
t_{i\to j}^r \quad \text{minimum in vehicle travel time between (i) and (j)}
D_{\text{min}} \quad \text{minimum passengers demand within service area to be covered directly (without transfer)}
D_{\text{ol}}^{\text{min}} \quad \text{minimum total passengers demand within service area to be covered indirectly (with one transfer)}
D_{\text{tot}}^{\text{min}} \quad \text{minimum total passengers demand within service area to be covered by set of bus routes}
d(R) \quad \text{Bus routes directness indicator}
d(R)^{\text{max}} \quad \text{maximum allowable directness value for bus routes}
T \quad \text{Bus route (r) travel time}
n \quad \text{is the number of nodes on route (r) = (n_1, n_2, n_3, \ldots, n_n)}
T^{\text{max}} \quad \text{maximum allowable bus route travel time}
N \quad \text{the set of transit network nodes}
C_r & C_2 \quad \text{weighting factors}
TBF \quad \text{total bus fleet}
Q_r^{\text{max}} \quad \text{maximum flow on any link of route (r)}
V \quad \text{Bus vehicle capacity}
L \quad \text{Bus allowable load factor}
A \quad \text{the set of network arcs}
A_{i}^+ \quad \text{the set of network arcs leaving node i}
A_{i}^- \quad \text{the set of network arcs entering node i}
t_a \quad \text{time on arc a}
v_a \quad \text{flow on arc a}
\omega_i \quad \text{total waiting time at node i}
x_a \quad \text{dummy variable, (1) if arc (a) belongs to attractive line set, (0) otherwise}
b_i \quad \text{flow generated at i}
D_i \quad \text{total flow go through node i}

a) Demand coverage classification: The demand for each node pair (d_{i,j}) is classified as -0- transfer, -1- transfer, -2- transfer or unsatisfied demand. Node pair demand d_{i,j} is considered to be satisfied directly (i.e. 0-transfer demand), if there is, at least, one bus route traversed both node (i) and (j). d_{i,j} is considered to be satisfied with -1- transfer, if , at least, one bus route traversed node (i) intersected with other route traversed node (j). d_{i,j} is considered to be satisfied with -2- transfer, if , at least, there is one bus
route intersected with route traverse (i) and route traverse (j). Remaining node pairs demand are considered unsatisfied demand. In Fig. 1, it is obvious that, \( d_{1,2} \) is \( 0 \)-transfer demand, \( d_{1,3} \) is \( 1 \)-transfer demand and \( d_{1,5} \) is \( 2 \)-transfer demand.

\[ d(R) = \frac{\sum_i (\sum_j d_{i,j} \times \frac{d_{i,j}}{d_{i,j}})}{\sum_i (\sum_j d_{i,j})} \]  

(1)

where; \( \sum_i (\sum_j d_{i,j}) \) represents network Total Demand.

a) **Transit route length**: it is measured, either in kilometers or in minutes. It is more appealing and practical to deal with routes length in time units. There is a constraint on maximum transit route length based on either heuristic guidelines, past experience or common practice accepted by transit planners. Longer bus routes may cause bus driver fatigue and consequently result in safety hazards. Maximum round trip should not exceed 2 hours [5].

b) **Maximum link flow**: After passenger assignment on the transit network, each bus route (r) would have a maximum link flow (\( Q^R_{\text{max}} \)). It is considered the ruler in setting bus route frequency and bus size to accommodate this maximum flow. Since in large networks, it is too difficult to keep track of all link flows. For each bus route, \( Q^R_{\text{max}} \) could be approximately obtained by multiplying the route’s directly satisfied flow with the flow factors;

\[ Q^R_{\text{max}} = (L + f_g) \times f_y \times D^R_0 \]  

(2)

\( f_f \) is transferring flow factor which accounts for the transferring flow on the route, \( f_y \) is maximum link flow fraction and it would be estimated as flows;

\[ f_f = \frac{n}{2(n-1)} \]  

(3)

\[ f_f = \frac{n+1}{2n} \]  

(4)

Where; \( D^R_0 \) is direct demand satisfied by route (r) and n is the number of nodes of bus route [6].

c) **Service Frequency**: it is simply referred to the number of buses running on a certain route per hour. The most commonly used service frequencies in the transit industry can be grouped into three categories; supply frequency, policy frequency, and demand frequency. Supply frequency is dependent on the operator’s resources including limited fleet size. It is the maximum frequency that the operator can provide under current resource and economic constraints. Demand frequency is determined by transit demand. This frequency is the minimum frequency that provides just enough capacity to meet the demand on the maximum link flow so that on the other links of this route, the demand is always less than the capacity. Policy frequency can serve as a lower bound and an upper bound for service frequency. It reflects transit network operation constraints and is usually used by transit operators when the demand is too high or too low. In the real world, as well as in the bus transit route network design process, the demand frequency approach is preferred because it reflects the purpose of transit operations, which is to provide customer- oriented service. Furthermore, the maximum bus route frequency and the minimum bus route frequency have been chosen as 30 buses/hr and 6 buses/hr [7, 8].

d) **Bus route capacity**: its concept differs from the ordinary concept of street-link capacity. The route capacity is based on the idea that the route service frequency shouldn’t exceed a predefined value and buses have a limit seating capacity, thus the maximum allowable flow on bus route-link;

\[ X_s \leq LF \times f_{\text{max}} \times V_i \]  

(5)

Where; \( X_s \) is the link passenger flow.

e) **Transfer penalty**: it is the term which reflects transit users inconvenience to make transfer. Transfer penalty is measured by equivalent in-vehicle time units. Penalty values are likely to vary across different geographic areas and change with demographics, socioeconomics, topography, clim- ate, quality of transfer facilities and etc.

f) **User Cost**: Total user cost can be easily expressed by the following expression;

\[ UC_k = \rho_v(TAT) + \rho_w(TWT) + \rho_s(TBT+TIVT) \]  

(6)

Where the terms in (6) represent respectively; users costs associated with mode (k) of travel, total passengers access time in minutes, total passengers waiting time in minutes, total passengers boarding time in minutes, total passengers in vehicle time in minutes. Whereas \( \rho_v, \rho_w, \rho_s \) are weighting values of access, waiting and in-vehicle time.

g) **Operator cost**: the main cost \( (OC_k) \) is the total bus fleet required for operating. Vehicles operating costs for vehicle type (k) are taken as the sum of the direct costs plus the indirect costs for the given input parameters such as:

The direct costs; (DC) are considered as three factors:

1. Costs of direct travelled distance (L) in kilometer; (CK).
2. Costs associated with time spent at bus stops; (CS).
3. Costs associated with personnel costs; (CP).
Indirect costs: (CF) were found in other studies to be nearly 12% of the direct costs these costs include maintaining buses license, insurance …etc [9].

Up to now, there is no general and solvable (in the same time) mathematical formulation for the TNDP; the reason is the high degree of problem complexity. There are five main sources of complexity that often preclude finding a unique optimal solution for TNDP, namely; problem formulation, non-linearity and non-convexity, combinatorial complexity, NP-hard and multi-objective nature of TNDP [4, 10-12]. However, it can be derived a general mathematical formulation to clarify TNDP objectives;

\[
\min \; C_1 \sum_{i \in \mathcal{E}} \sum_{j \in \mathcal{N}} d_{ij} \times t_{ij} + C_2 \sum_{r \in \mathcal{R}} f_r \times T
\]

s.t.

\[
D_{\text{low}}(R) \leq D_{\text{max}} \leq D_{\text{high}} \leq D_{\text{min}} (8)
\]

\[
D_{\text{out}}(R) = D_{\text{in}} + D_{\text{out}} \geq D_{\text{in}} (9)
\]

\[
d(R) = \sum_{i \in \mathcal{N}} d_{ij} \times \frac{t_{ij}}{v_{ij}} \leq d_{\text{max}} (11)
\]

\[
T = \sum_{t=1}^{\text{max}} \sum_{j \in \mathcal{N}} t_{ij} \leq T_{\text{max}} (12)
\]

\[
\sum_{r \in \mathcal{R}} f_r \times T \leq T_{\text{BF}} (13)
\]

Equation (7) represents the general objectives of TNDP. First term concerns with users costs. Second term concerns with operator’s cost. C1 & C2 weighting factors reflect the relative importance of two cost components. By varying C1 & C2, one can generate different trade-offs between users and operator costs. (8 - 10) represent demand coverage constraints. Inequality (11) represents route directness. (12-14) reflect the operational parameters.

2. STATE OF ART

TNDP is a complete hierarchical multi-disciplinary process, see Fig. 2. It shows the general framework of the TNDP. The Main data are the demand trips between different traffic zones of the city and the street network structure. Passengers’ demand for travel among Origin – Destination (O/D) locations represents the need of transport system. Each zone has associated with one zone centroid and group of passengers with their desired (O/D) trips which are concentrated at zone centroid. Transit modes supply services are routed in an environment to pick – up and deliver passengers demand among – what is called – Transit (O/D) locations through selection of efficient bus routes from a given street network representation.

For certain zone, the number of O-D trips is a product of population, employment, land-use and transit (expected) level of service in this zone. Demand is an essential element for the TNDP. The more precise transit O/D demand estimation, the more adequate TNDP solution is obtained. There are two approaches for transit (O/D) matrix estimation. First approach is to use the traditional four-step travel forecasting model. This approach is used for networks that would be designed from scratch (new towns without any existing transit systems). Second approach, is to use counting data at existing bus stops. Although, literature review concerning auto O-D estimation is extensive [13, 14], there are few works concerning transit O-D estimation [15].

Previously, two approaches for transit demand are tackled. The first approach considers demand fixed and independent of service quality. Whereas the second approach considers demand variable and dependent on simultaneous distribution – modal split. However, the variable demand assumption is more appealing, the first assumption is common used for the purpose of simplicity [16, 17]. Two basic levels of the street network representation could be classified for The TNDP: First type is zone centroid level where each zone demand is aggregated at its centroid. Second level is node level where each node in the network is considered as a potential bus stop (origin). It isn’t practical solution to consider all these nodes as bus stops, so a filtering stage should be provided to select candidate bus stops (nodes) according to each node demand and maximum allowable walking distance [18, 19]. Route design stage is the essential part of The TNDP. It constitutes the strategic to tactical planning term in the design process. It forms the combinatorial complexity of the problem [20 - 25]. Route connectivity represents the major obstacle in mathematical formulation for the TNDP [26]. Three basic approaches for route construction would be defined, namely; graph based methods, mathematical based methods and greedy based methods.

Graph based methods use the network graph representation to generate routes according to shortest paths or k-shortest paths strategies [27, 28]. Mathetical based methods use the techniques found in Travel Salesman Problem (TSP). TSP is one of the most widely studied combinatorial optimization problems. However, its statement is deceptively simple: a salesman seeking the shortest tour through (N) clients (cities) that allows each client to be visited once and only once [29- 31]. It is used in some researches to investigate its applicability for bus routes generation [32-34]. Greedy based methods use a node as start node, then they begin to construct route by connecting a node adjacent to the considered node. This process is repeated, until the route construction is completed, as shown in Fig 3. The next node to visit is selected according to a strategy [35-37].
Transit passengers in many cases have to deal with overlapping bus route with some routes sharing sections and common stops. This problem is a sub-problem of transit passenger assignment, called common-lines problem. Various assumptions and studies are made in order to track passengers’ behavior towards a given supply of transit service [45–49].

The concept of clever passenger may be considered as the basis of all transit assignment models. Passengers would minimize the sum of waiting time and in-vehicle time in their boarding strategies. If there is more than one route serving an origin node (i) and destination node (j). This would lead the passengers, who wish to travel from (i) to (j), to determine a sub-set of bus routes (attractive lines) boarding the first incoming bus of these routes. Mathematically, it could be formulated as follows:

$$\begin{align*}
\text{minimize} & \quad \sum_{(a,E) \in T} a_{ij} + \sum_{i \in N} o_i \\
\text{s.t.} & \quad \sum_{(a,E) \in T} v_a - \sum_{a \in E} v_a = b_i \quad \forall \ i \in N \\
& \quad o_i = \frac{d_i}{\sum_{a \in A_i} f_a s_a} \quad \forall \ i \in N \\
& \quad v_a \geq 0 \quad \forall \ a \in A_i
\end{align*}$$

(15)

The model in (15) states that equilibrium occurs when all users succeed in minimizing both travel and expected waiting time. It is worth noting; that expected waiting time is equal to the inverse of line(s) frequency. (15-18) are mixed integer non-linear model which is difficult (impossible) to be solved for large scale networks. Other modeling difficulties are three fold; forcing the number of allowed transfer, incorporating the transfer penalty and congestion modeling.

In the stated model, users may do unlimited number of transfers opposing to the reality. Incorporating transfer penalty leads to expand the network to incorporate transfer links associated with transfer penalty value. Congestion refers to the potentiality that user may not board the first incoming bus (of attractive lines) due to congestion. Many studies have attempted to overcome these aspects [18, 42, 50, 51].

In [6, 11], they developed Network Analysis Procedure (NETAP) to assign passengers through iterative procedure. It alleviated the complexity of the mathematical model; however it has no theoretical prove to reach a convergence state and many simplifications are made. There are limited works on heuristics transit assignment models like [36, 40].

3. CONCLUSION

In this paper, the state of art of the TNPD is reviewed to expand the knowledge of the problem issues and challenges among recent researchers. It is obvious that there are still suspended items which need extensive studies. Route construction techniques and selection procedures are still widely open for further studies. Candidate bus stops problem is received little attention in the literature review. It needs more investigation about the way of its incorporation in the TNPD design process. Accurate and dynamic transit O/D predictors are also worth more investigation. The behavior of users who would transfer needs more clarification and modeling. In general terms, a global solution methodology for the TNPD is a challenging task besides none of the review studies have claimed reaching the optimal solution of the problem.
4. ACKNOWLEDGMENT
I would like to thank Eng. Marwa Abd El-Gwad for her valuable assistant and support.

5. REFERENCES
[6] Shih, M.-C. and H.Mahmassani, A design methodology for bus transit networks with coordinated operation, in SWUTC/94/60016-1, Center for Transportation, Bureau of Engineering Research. 1994, the University of Texas at Austin, Austin, Texas.


