https://doi.org/10.1016/j.trb.2018.03.007

Optimal design of intersecting bimodal transit networks in a grid city

Wenbo Fan[a,b], Yu Mei[b], Weihua Gu[b*]

a. Department of Traffic Engineering, Southwest Jiaotong University, Chengdu, China

b. Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China

Abstract

The urban transit system in a real city usually has two major components: a sparse express (e.g. rail) network and a dense local (e.g. bus) network. The two networks intersect and interweave with each other throughout the city to furnish various route options for serving transit patrons with distinct ODs. The optimal design problem of this bimodal transit system, however, has not been well explored in the literature, partly due to the difficulty of modeling the patrons' complex route choice behavior in the bimodal networks. In light of this, we formulate parsimonious continuum models for minimizing the total cost of the patrons and the transit agency for an intersecting bimodal transit network in a grid city, where the perpendicular local and express lines intersect at transfer stops. Seven distinct route types are identified in this network, which represent realistic intra- and inter-modal route options. A lower-level assignment problem between these routes is embedded in the upper-level network design optimization problem. We develop an efficient method to find near-optimal designs of the intersecting network. Numerical results unveil a number of insightful findings, e.g., that sizable cost savings are observed for the intersecting bimodal design as compared to the single-mode designs for moderate to high demand levels, and that only moderate benefits are observed as compared to the trunk-feeder designs under certain operating conditions. We also show that the conventional practice of designing the local and express networks separately would greatly undermine the benefit of the bimodal system.

Key words: bimodal transit network; continuum models; optimal design; transit route choice

1. Introduction

An urban public transit system often consists of two overlapping and interweaving single-mode networks: a local bus network that features high line and stop densities but low speed and operating costs, and an express transit network that features high speed and capacity, but has to be sparsely spaced due to the high costs. The latter is often operated by Bus Rapid

* Corresponding author. Tel.: + 852 27666189; fax: + 852 23301544.
E-mail address: weihua.gu@polyu.edu.hk.

Transit (BRT) or rail. Cities having this kind of bimodal transit systems include Beijing, London, San Francisco (bus plus metro), and Bogota (bus plus BRT), among many others. The interweaving local and express networks furnish multiple route options, so that patrons with distinct OD can choose the routes that best suit their needs. For example, short-distance travelers whose origins and destinations are far from the express lines can choose to travel by local lines only, and long-distance travelers can take the local service as feeder to access the express lines. Survey data in real cities served by bimodal transit systems have confirmed that there are significant proportions of patrons who chose more than one route type. For example, it was reported for a number of major cities in the world that the percentage of transit trips involving intermodal transfers was among 25-30%, while the rest of the trips are local-only or express-only (Guo, 2008).

Despite the wide presence of bimodal transit systems, a fundamental question has yet to be answered: does this interweaving and redundant bimodal transit network, if optimally designed, perform better than other network designs (e.g. single-mode networks) under certain operating conditions? There seems to be a chance for the bimodal network to win, thanks to the potential patron benefits resulting from the multiple route options. Unfortunately, at present there is no study that examines how the patrons' benefits trade off optimally with the extra agency cost for providing the service redundancy, and how the optimal system design performs in various operating environments, as we shall see next.

1.1 Literature review and rundown of the paper

Studies on transit network optimization can be classified into two categories: those on discrete models and those on continuum models. Discrete models (see the detailed reviews by Kepaptsoglou and Karlaftis, 2009, and Ibarra-Rojas et al., 2015) are often built upon a well-defined graph with nodes and links representing the underlying street network, the discrete demand points, and candidate stops. These models are thus able to account for the specific OD and irregular street layouts in real cities. However, the resulting formulation (typically a vehicle routing problem) is usually NP-hard, and cannot be solved to global optimum by exact solution methods (e.g. branch-and-cut), save for a few networks of very small sizes (e.g. Guan et al., 2003; Wan and Lo, 2003). On the other hand, heuristic solution methods (e.g. genetic algorithm) were often used to provide reasonably good solutions to small-to-medium-scale problems with few dozens of nodes and links (Ceder and Wilson, 1986; Fan and Machemehl, 2004). For large-scale networks, especially those with multiple transit modes (e.g. Wan and Lo, 2009), the heuristic methods either cannot guarantee the solution quality (i.e. how close the heuristic solution is to the global optimum), or have intolerable computation times.

On the other hand, continuum models (e.g. Wirasinghe, 1980; Daganzo, 2010a; Chen et al., 2015; Gu et al., 2016) are often built upon idealized network forms (e.g. grid, radial, ring-radial) and demand patterns (e.g. uniformly or concentrically distributed demand). Despite the idealization, the continuum models are parsimonious and can usually be solved to global optimality or near optimality via computationally efficient numerical methods. Therefore, the continuum models are often used to examine the optimal network designs for a wide range of operating conditions, and to explore general insights into the cause-and-effect relations between key parameters and the optimal design. Thus, this approach is ideal for examining the fundamental question described above.

Regrettably, most studies on continuum models of transit system design are focused on single-mode networks (Newell, 1973; 1979; Holryod, 1967; Vuchic and Newell, 1968; Byrne, 1975; Wirasinghe and Ghoneim, 1981; Chien and Scholfeld, 1997; Wirasinghe, 2003; Daganzo 2010a, b; Estrada et al., 2011; Badia et al., 2014; Ouyang et al., 2014). Some works examined local-and-express systems (Gu et al., 2016; Fan et al., 2017) or other systems with differentiated transit services (Chang and Schonfeld, 1993; Li and Quadrifoglio, 2011; Freyss et al., 2013; Gu et al., 2016). In the above-cited works, however, the differentiated services were still operated by a single transit mode, and they were designed to serve a corridor only.

To our best knowledge, the only type of multimodal transit systems that has been studied in the literature of continuum models is the trunk-feeder system, where buses serve as feeders to carry patrons to or from the trunk (e.g. rail) lines. Among the limited number of works on this topic, however, most are about corridor service designs (Wirasinghe et al., 1977; 1980; Hurdle and Wirasinghe, 1980; Chien and Schonfeld, 1998; Sivakumaran et al., 2012; Fan and Mei, 2018), while Sivakumaran et al. (2014) seems to be the only one that modeled a city-wide trunk-feeder network[1]. This reference compared optimal trunk-feeder networks (bus feeding BRT and bus feeding rail) against the optimal single-mode networks (bus-only, BRT-only, and rail-only), and found the former triumphed over the latter in a wide

[1] Here our discussion is limited to the networks with *fixed-route* feeder lines. Trunk-feeder networks with flex-route feeders (e.g. Chen and Nie, 2017a, b, 2018) are out of our scope.

range of operating conditions. However, in this trunk-feeder network each patron has to transfer three times in any trip, which greatly undermines the practicality of this system. Note in real urban transit systems that the average number of transfers per trip is only around 1.5 (APTA, 2007). The network structure in Sivakumaran et al. (2014) also penalizes short trips: they still have to transfer three times between the feeder buses and trunk lines since no direct local route is offered.

Also note that most studies on continuum modeling of transit networks (including the studies on multimodal systems; e.g. Sivakumaran et al., 2014) have assumed that each patron has only one route option. This simplification of patrons' route choice might be necessary for reducing the complexity of the modeling work, however it is unrealistic since most trips in a real bimodal urban transit system are offered multiple route options (APTA, 2007). The only exception seems to be Saidi et al. (2016), which modeled a relatively simple, single-mode transit network (consisting of one ring line and multiple radial lines). The route choice model is still too simple to represent patrons' realistic route choice behavior in a more complex bimodal network.

In light of the above, we formulate continuum models for optimizing a specific type of bimodal transit network atop a generic city with grid street pattern. We term this network as the "intersecting bimodal network", in which express and local lines intersect each other at the transfer stops (see Figure 1). Although being special and idealized, this network furnishes a number of route options, including express-only, local-only, and intermodal routes that involve 1-3 transfers. Particularly, the local lines in the network can serve both the short trips directly and the longer trips as feeder service to/from the express stations. This network structure is also flexible, since it allows the ratio between the numbers of express and local lines to vary. In extreme cases the network will reduce to a single-mode one. These idealized features are sufficient for representing realistic bimodal transit systems for the high-level planning purpose. The network layout is described in detail in Section 2.1.

Our models are presented in Sections 2.2-2.5, which minimize the generalized cost for this intersecting transit network, including the patrons' travel cost and the transit agency's capital and operating costs (Daganzo, 2010a, b). The models explicitly account for patrons' route choice among various options (see Section 2.3 for the route assignment model). The optimization is solved by a bi-level approach, where the upper level optimizes the design variables (line and stop spacings, headways, etc.), and the lower level finds the route assignment equilibrium under a given design. This solution approach is described in Section 3.

Section 4 of the paper examines numerical cases under a wide range of operating parameters. The results show that the intersecting bimodal network can often outperform single-mode networks. This manifests the benefit of providing service redundancy under certain conditions. Final remarks and potential extensions are discussed in Section 5.

2. Methodology

The layout of an intersecting bimodal network is described in Section 2.1. The transit system's generalized cost optimization problem is formulated in Section 2.2. The detailed cost models for the transit patrons and the agency are developed in Sections 2.3 and 2.4, respectively. Finally, the critical vehicle occupancies for the express and local lines, which are used in the vehicle capacity constraints of the problem formulation, are formulated in Section 2.5. The notation used in this paper is summarized in Appendix A.

2.1 The intersecting express-local network

The intersecting network is illustrated in Figure 1, where the express lines (the thicker ones in the figure) form a grid of spacing S_1, and the local lines (the thinner ones) are distributed evenly between any two neighboring parallel express lines with a spacing of S_2. We have $S_1 = mS_2$ where $m \geq 1$ is an integer. The express lines intersect at the transfer stops marked by the solid squares in the figure, which we term as Type-1 stops. An East-West (E-W) express line intersects with a North-South (N-S) local line at a transfer stop marked by a solid dot, which we term as a Type-2 stop; similarly, a N-S express line intersects with an E-W local line at a Type-3 stop, which is also marked by a solid dot in the figure. Also, the local lines cross each other at the Type-4 stops marked by the hollow circles. Finally, intermediate, non-transfer stops are evenly distributed on each local line with a spacing of S_3, as shown by the right side of Figure 1. Note that if $m = 1$, the network reduces to a single-mode network with express lines only.

To simplify the modeling work, we adopt the following assumptions:

A1. The city is *edgeless* and has a dense grid street network. This assumption eliminates any unnecessary irregularities in the mathematical models that may occur at the city boundaries.

A2. The demand is exogenous and inelastic. The trip origins are uniformly distributed over the city with density λ patrons/km^2/hour. For a given trip origin, the destinations are uniformly distributed inside a square of $2L \times 2L$ centered at the origin. As such, the distance that a trip covers in each of E-W and N-S directions follows a uniform distribution over $[0, L]$. Thus, the average trip length is L. The uniform demand assumption has been used by a number of studies on transit network optimization (e.g., Daganzo, 2010a; Chen et al., 2015) to obtain useful generic insights.

A3. Vehicles used in each mode (express or local) have the same speed, passenger-carrying capacity, and cost structure.

A4. A flat fare per trip is applied to the entire transit system, regardless of a patron's choice between express and local lines, the travel distance, and the number of transfers involved. Under this assumption, the fare will not affect the patrons' route choice in the transit network.

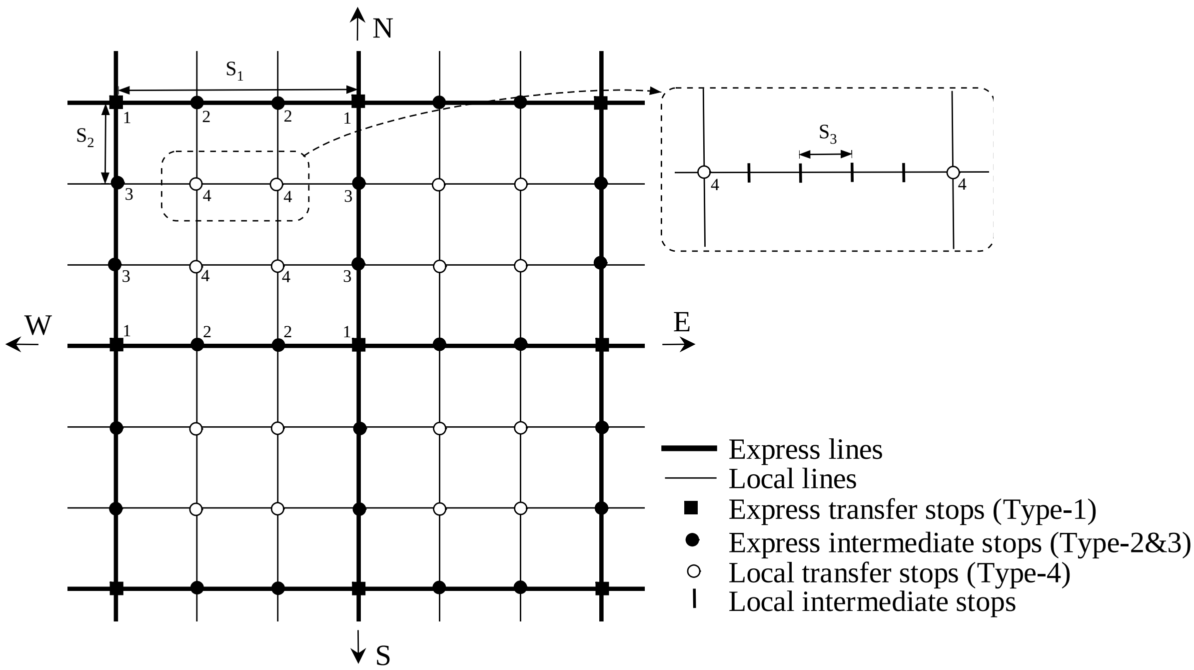


Figure 1. Structure of an intersecting express-local transit network

2.2 Mathematical formulation

The minimization of the system's overall generalized cost, Z, is formulated as the following nonlinear program:

$$\min_{S_1, S_2, S_3, H_{lo}, H_{ex}} Z = T + AC \tag{1a}$$

subject to:

$$O_{lo} \leq C_{lo} \tag{1b}$$

$$O_{ex} \leq C_{ex} \tag{1c}$$

$$H_{lo} \geq H_{lo}^{min} \tag{1d}$$

$$H_{ex} \geq H_{ex}^{min} \tag{1e}$$

$$0 \leq S_1 = mS_2 \leq \frac{L}{m_0} \tag{1f}$$

$$S_2 = m' S_3 \tag{1g}$$

$$m, m' \in \{1,2, \dots\} \tag{1h}$$

where H_{lo} and H_{ex} denote the local and express service headways, respectively; T and AC denote the average patron's travel cost and the average agency cost per trip, both in the unit of time. Constraints (1b-c) ensure that the maximum numbers of passengers aboard a local and an express vehicle, denoted by O_{lo} and O_{ex} respectively, do not exceed the vehicles' passenger-carrying capacities, C_{lo} and C_{ex}. Constraints (1d-e) specify the minimum feasible headways for the local and express lines, H_{lo}^{min} and H_{ex}^{min} respectively, which are determined by the safety requirements or vehicle-carrying capacities of the transit lines (see e.g., Gu et al., 2016). Finally, constraints (1f-h) specify the feasible ranges of the stop spacing variables, and that the ratios $\frac{S_1}{S_2}$ and $\frac{S_2}{S_3}$ must be integers. Constraint (1f) also stipulates that S_1 is no greater than $\frac{L}{m_0}$ for a given constant $m_0 \geq 1$. This constraint ensures that $L \gg S_1$ is a valid assumption.

2.3 Patron's average travel cost

A patron's trip cost consists of the access and egress cost to/from the transit system (by walking), the waiting times at the origin and transfer stops, the in-vehicle travel time, and a transfer penalty (Daganzo, 2010a; Chen et al., 2015; Gu et al., 2016). When multiple route options are available, the patron will choose the minimum-cost route according to Wardrop's user equilibrium principle (Wardrop, 1952). Thus, in this section we first enumerate the route options for various patrons, grouped by their OD (Section 2.3.1), and then develop the trip cost models for each route option (Section 2.3.2). Finally, the demand is assigned to the route options according to the user equilibrium, and thence the average trip cost is derived (Section 2.3.3). To simplify the route assignment model, we further make the following assumptions:

A5. An optional route for a patron must satisfy the following conditions: i) at either end of her trip, the patron will only choose between two lines, regardless of whether they are express or local: the nearest E-W line and the nearest N-S line; ii) the route must contain no more than 3 transfers; and iii) the patron will access and egress the transit system through the nearest stops *on the selected lines*. This assumption rules out some uncommon route options, e.g. routes with more than 3 transfers.

A6. If the patron has more than one route options, she will choose the route with the minimum patron's cost.[2]

2.3.1 Demand groups and their route choice options

Under assumption A5, a patron's accessible transit lines depend on her origin and destination. For illustration, Figure 2 shows a squared area enclosed by two express and two local lines. This area is divided into four equal-sized zones marked by circled numbers 1, 2, 3, and 4. Patrons originated in zone 1 can access two perpendicular express lines through the Type-1 stop; patrons in zones 2 and 3 can access one express line through the Type-2 or 3 stop, and one local line through the nearest local stop; and patrons in zone 4 can access two perpendicular local lines through the nearest local stop on each line. We term the trip origins in zone i a Type-i origin.

The four origin types can be applied to the entire intersecting network; i.e., a trip origin whose nearest *transfer* stop is a Type-i stop ($i \in \{1,2,3,4\}$) is a Type-i origin. The trip destinations are similarly classified by the same four types. Hence, the trip ODs can be classified into ten types: $\{(1,1),(1,2),(1,3),(1,4),(2,2),(2,3),(2,4),(3,3),(3,4),(4,4)\}$. Here we consider OD pairs (i,j) and (j,i) are of the same type, $\forall i,j \in \{1,2,3,4\}$, because the route options of (j,i) are simply the reverse of those of (i,j). These ten types of OD pairs can be further divided into five groups, such that the OD pairs belonging to each group have the same set of route options, and their route assignment can be described by the same mathematical equations. The five OD groups, denoted by G_k $(k = 1,2,3,4,5)$, and their associated sets of route options, denoted by R_k $(k = 1,2,3,4,5)$, are summarized in Table 1.

[2] We choose this deterministic route assignment model to simplify the modeling work, albeit stochastic route assignment models (e.g., a logit model) can also be incorporated. We believe this will not compromise the major findings drawn from the models.

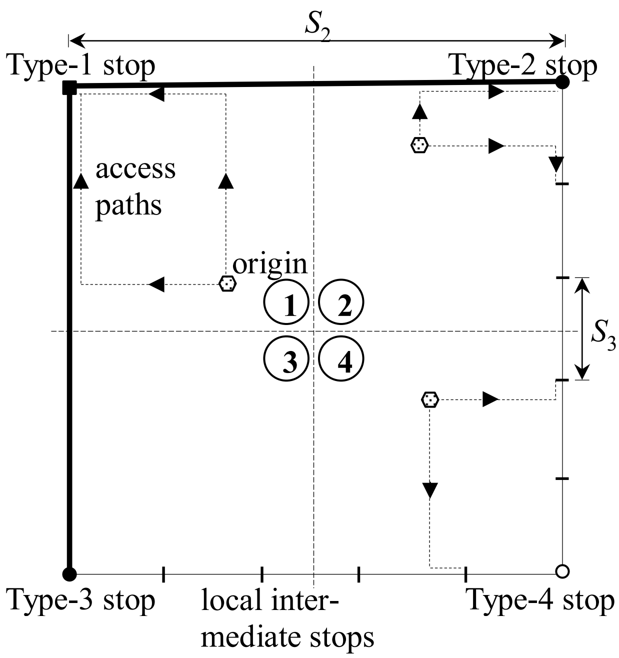


Figure 2. Trip origin zones and their access paths

For each OD group in Table 1, we find the route options that satisfy assumption A5. To further reduce the number of route options, we delete those apparently inferior ones; e.g. the "local-local" route for OD type (2,3). We also specify that a patron is willing to walk to a further line if that could save a transfer (Guo and Wilson, 2004). For example, the "local-local-local" route for OD type (4,4) is considered inferior to the "local-local" route and thus deleted, although the former has a slightly smaller access cost; see Figure 3. The remaining route options for the five OD groups are illustrated in Figures 4a-e, respectively. Each arrowed solid line in the figures represents a route option that serves a specific OD type (for brevity we only showed transfer stops in the figure). Specifically, all the OD types in G_1 will be served by route ee (express-express) only (Figure 4a). While other feasible route options (e.g. local-express) may exist, they are certainly inferior to route ee. For OD type (1,4) in G_2, two route options exist: le (local-express) or el, and lee (local-express-express) or eel (Figure 4b). Note that in route lee, the local line serves as a feeder to the express lines. The same set of route options apply to the OD types in G_3 (Figure 4c). (We separate these OD types into two group, G_2 and G_3, because the route assignment models are different between the two groups; see Section 2.3.3.) Similarly, patrons in G_4 have two route options: le/el and eee (express-express-express), as illustrated in Figure 4d. Patrons in G_5 have three route options: ll (local-local), lel (local-express-local), and $leel$ (local-express-express-local); see Figure 4e. In route option $leel$, both local segments function as feeders. In the following discussions, we abbreviate "le/el" to "le", and "lee/eel" to "lee" for simplicity.

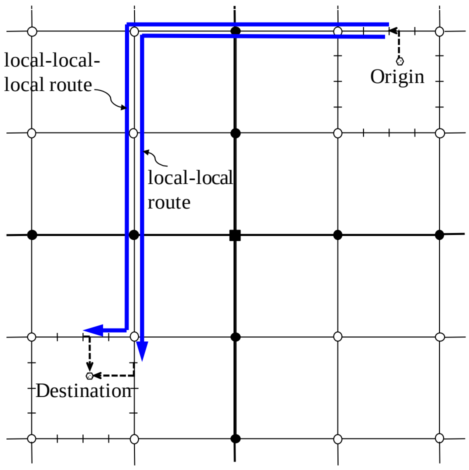
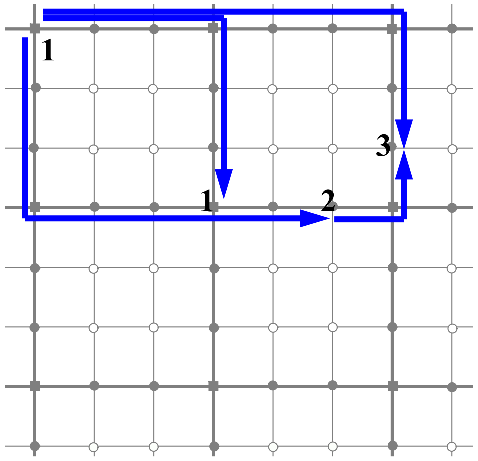


Figure 3. Comparison of a local-local-local route with a local-local route

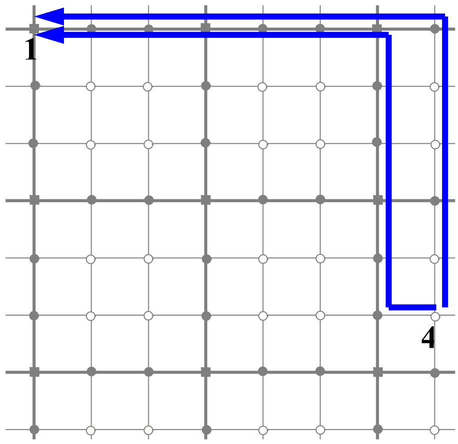
Table 1. Transit route options for different groups of OD pairs

OD group	Set of route options
$G_1 = \{(1,1), (1,2), (1,3), (2,3)\}$	$R_1 = \{ee\}$
$G_2 = \{(1,4)\}$	$R_2 = \{le/el,\ lee/eel\}$
$G_3 = \{(2,4), (3,4)\}$	$R_3 = \{le/el,\ lee/eel\}$
$G_4 = \{(2,2), (3,3)\}$	$R_4 = \{le/el,\ eee\}$
$G_5 = \{(4,4)\}$	$R_5 = \{ll,\ lel,\ leel\}$

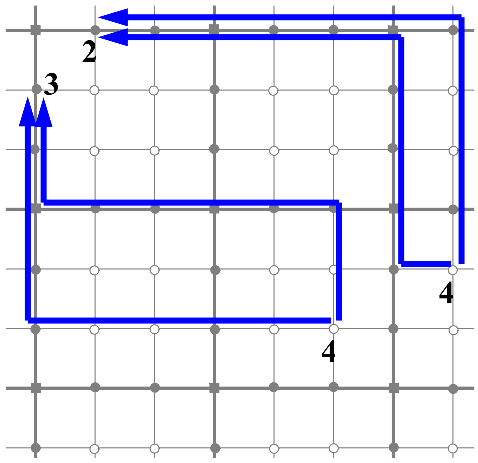
Note that some route options have more than one equivalent routes; e.g., there exist two equivalent routes of type ee for serving an OD of type $(1,1)$, and more routes of type eee for serving an OD of type $(2,2)$ or $(3,3)$. These equivalent routes have the same travel cost. Here we assume that a patron has an equal chance to take any of these equivalent routes.



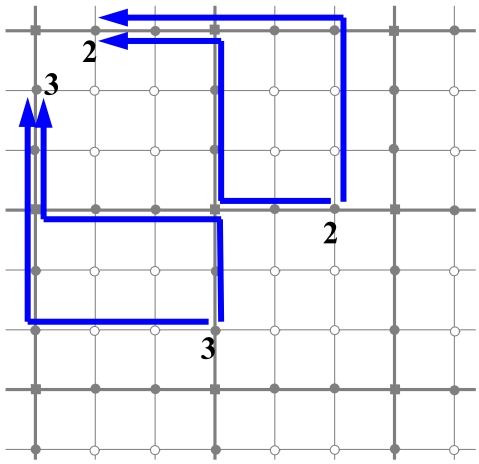
a. Route ee of OD group G_1



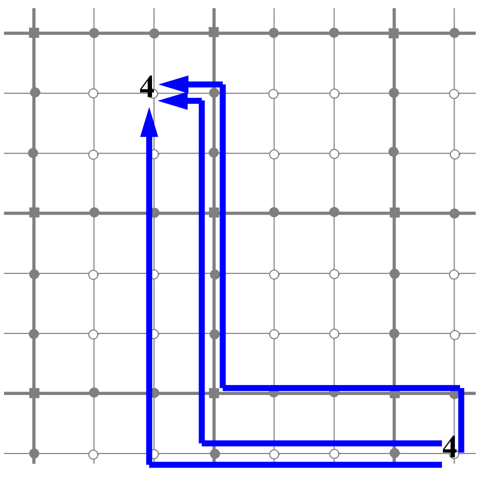
b. Routes le/el and lee/eel of OD group G_2



c. Routes le/el and lee/eel of OD group G_3



d. Routes le/el and eee of OD group G_4



e. Routes ll, lel and $leel$ of OD group G_5

Figure 4. Illustration of the route options associated with five OD groups

2.3.2 Route travel cost models

We consider an arbitrary patron whose *direct* travel distance is X in the E-W direction and Y in the N-S direction. The travel cost, t_I through a route of type $I \in \bigcup_{k=1}^{5} R_k = \{ee, le, lee, eee, ll, lel, leel\}$, can be formulated as the sum of four components:

$$t_I = t_I^A + t_I^W + t_I^T + t_I^E, I \in \bigcup_{k=1}^{5} R_k \tag{2}$$

where t_I^A denotes the patron's access and egress time by walking; t_I^W the total waiting time at the origin and transfer stops; t_I^T the total in-vehicle travel time; and t_I^E the transfer penalty.

For simplicity, we approximate t_I^A by the *average* access and egress time for all the patrons who may choose route option I. For example, t_{ee}^A is approximated by $\frac{S_2}{V_w}$ (see Figure 2), where V_w is a patron's walking speed. We also approximate the patron's waiting time at any origin or transfer stop by half of the headway of the transit vehicle to board (Daganzo, 2010b). We further approximate the in-vehicle travel distance of a local *feeder* segment by the *average* distance of that segment (e.g. $\frac{S_1}{4}$ if the direction of the feeder segment is given). We believe the above approximations have only small impacts on the calculation of the t_I's, and thus on the patrons' route choices. With these assumptions, the t_I for $I \in \bigcup_{k=1}^{5} R_k$ can be formulated as functions of X and Y only, which are presented as follows:

$$t_{ee} = \frac{S_2}{V_w} + H_{ex} + \frac{X+Y}{V_{ex}} + \xi_{ee} \tag{3}$$

$$t_{le} = \frac{3S_2+S_3}{4V_w} + \frac{H_{lo}+H_{ex}}{2} + \left(\frac{d_{lo}}{V_{lo}} + \frac{d_{ex}}{V_{ex}}\right) + \xi_{le} \tag{4}$$

$$t_{lee} = \frac{3S_2+S_3}{4V_w} + \left(\frac{H_{lo}}{2} + H_{ex}\right) + \left(\frac{S_1}{4V_{lo}} + \frac{X+Y}{V_{ex}}\right) + (\xi_{le} + \xi_{ee}) \tag{5}$$

$$t_{eee} = \frac{S_2}{V_w} + 3\frac{H_{ex}}{2} + \frac{X+Y}{V_{ex}} + 2\xi_{ee} \tag{6}$$

$$t_{ll} = \frac{5S_2+6S_3}{12V_w} + H_{lo} + \frac{X+Y}{V_{lo}} + \xi_{ll} \tag{7}$$

$$t_{lel} = \frac{S_2+S_3}{2V_w} + \left(\frac{H_{ex}}{2} + H_{lo}\right) + \left(\frac{\min(X,Y)}{V_{lo}} + \frac{\max(X,Y)}{V_{ex}}\right) + 2\xi_{le} \tag{8}$$

$$t_{leel} = \frac{S_2+S_3}{2V_w} + (H_{ex} + H_{lo}) + \left(\frac{5S_1}{12V_{lo}} + \frac{X+Y}{V_{ex}}\right) + (2\xi_{le} + \xi_{ee}) \tag{9}$$

The first term in the right-hand-side (RHS) of each equation of (3-9) is the average access and egress time for a given route type. Note from Figure 2 that the average access distance to a Type-1, 2 or 3 stop on an express line is $\frac{S_2}{2}$, and the average access distance to a local stop on a local line of *given direction* is $\frac{S_2+S_3}{4}$. Hence, the total access/egress distance is

S_2 if both ends of the route are at express stops; $\frac{S_2+S_3}{2}$ if both ends are at local stops of lines in a given direction; and $\frac{3S_2+S_3}{4}$ if one end is at an express stop and the other is at a local stop of a line in a given direction. This explains the access/egress cost terms in (3-6) and (8-9). For route type ll, the patron can freely choose the nearest local line on one end of her trip (hence the access distance is $\frac{S_2}{6}+\frac{S_3}{4}$), but on the other end the direction of the local line is fixed because it has to be perpendicular to the first local line (the access cost is $\frac{S_2+S_3}{4}$), thus the total is $\frac{5S_2+6S_3}{12}$. The second terms in the RHS of (3-9) are the average total waiting times at the origin and transfer stops.

The third term is the total in-vehicle travel time, where V_{lo} and V_{ex} are the commercial speeds of local and express transit vehicles, respectively. They are calculated as follows:

$$\frac{1}{V_{lo}}=\frac{1}{v_{lo}}+\frac{t_d^{lo}+t_b^{lo}}{S_3} \tag{10a}$$

$$\frac{1}{V_{ex}}=\frac{1}{v_{ex}}+\frac{t_d^{ex}+t_b^{ex}}{S_2} \tag{10b}$$

where v_{lo} and v_{ex} are the cruise speeds of local and express vehicles, respectively; t_d^{lo} and t_d^{ex} are the constant delays per stop due to acceleration and deceleration of local and express vehicles at a stop, respectively; t_b^{lo} and t_b^{ex} are the dwell times spent on loading and unloading patrons at a stop for local and express vehicles, respectively. The t_b^{lo} and t_b^{ex} are proportional to the number of boarding patrons at each stop. The models for computing these two variables are detailed in Appendix B.

For route types ee, eee, and ll, the in-vehicle travel distance is approximately $X+Y$; see Figures 4a, d, and e. For route types lee, the in-vehicle travel distance along the express lines is $X+Y$, and the local feeder segment is approximately $\frac{S_1}{4}$; see Figures 4b and c. Route type $leel$ is similar to lee, except that the two local feeder segments sum up to $\frac{5S_1}{12}$: the patron can freely choose the feeder direction on one trip end, but not on the other trip end. For route type lel, the patron will take the express line to travel through the longer segment between the E-W and N-S directions. Thus, the in-vehicle travel distance along the express line is $\max(X,Y)$, and the distance along the local lines is $\min(X,Y)$; see Figure 4e. For route type le, the local and express in-vehicle travel distances, denoted by d_{lo} and d_{ex} respectively, are formulated as follows (see Figures 4b, c, and d):

$$d_{lo} = \begin{cases} \min(X,Y), \text{for OD type } (1,4) \\ Y, \text{for OD types } (2,4) \text{ and } (2,2) \\ X, \text{for OD types } (3,4) \text{ and } (3,3) \end{cases} \tag{11a}$$

$$d_{ex} = \begin{cases} \max(X,Y), \text{for OD type } (1,4) \\ X, \text{for OD types } (2,4) \text{ and } (2,2) \\ Y, \text{for OD types } (3,4) \text{ and } (3,3) \end{cases} \tag{11b}$$

Finally, the last terms in the RHS of (3-9) indicate the transfer penalties, where ξ_{ee}, ξ_{ll}, and ξ_{le} denote the penalties for transfers within the express system, within the local system, and between the express and local systems, respectively.

2.3.3 Route assignment and a patron's average travel cost

The route assignment process is described in the following three steps:

Step 1: Determine the probability, P_{G_k}, that a trip belongs to OD group G_k $(k = 1, 2, \ldots, 5)$. We have $P_{G_k} = \sum_{(i,j)\in G_k} P_{(i,j)}$ $(k = 1,2,\ldots,5; i,j \in \{1,2,3,4\})$, where $P_{(i,j)}$ is the probability of OD type (i,j) as calculated by:

$$P_{(i,j)} = \begin{cases} p_i^2 & \text{if } i = j \\ 2p_i p_j & \text{if } i \neq j \end{cases}, i,j \in \{1,2,3,4\} \tag{12}$$

The p_i $(i = 1,2,3,4)$ denotes the probability that the trip origin or destination is of type i. By examining a typical "tile" of the intersecting network enclosed by four neighboring express lines (see Figure 5), the p_i's can be obtained as follows:

$$p_1 = \frac{S_2^2}{S_1^2} \tag{13a}$$

$$p_2 = p_3 = \frac{S_1 S_2 - S_2^2}{S_1^2} \tag{13b}$$

$$p_4 = \frac{(S_1 - S_2)^2}{S_1^2} \tag{13c}$$

Step 2: For each OD group G_k $(k = 1, 2, \ldots, 5)$, determine the probability, $Q_I^{G_k}$, that a patron chooses route option $I \in R_k$. We have:

$$Q_I^{G_k} \equiv \text{Prob}\left(t_I = \min_{J\in R_k}(t_J)\right), I \in R_k, k = 1,2,3,4,5 \tag{14}$$

For OD group G_1, we have $Q_{ee}^{G_1} = 1$. For the other OD groups, the formulas for $Q_I^{G_k}$ are derived in Appendix C, where the local and express commercial speeds V_{lo} and V_{ex}

are taken as given. In fact, V_{lo} and V_{ex} are functions of the numbers of boarding passengers at local and express stops, and are thus dependent on the route assignment of demand. Hence, the $Q_I^{G_k}$ ($I \in R_k, k = 1, 2, \ldots, 5$), V_{lo}, V_{ex}, t_b^{lo}, and t_b^{ex} will be updated in an iterative fashion until an equilibrium route assignment is achieved.

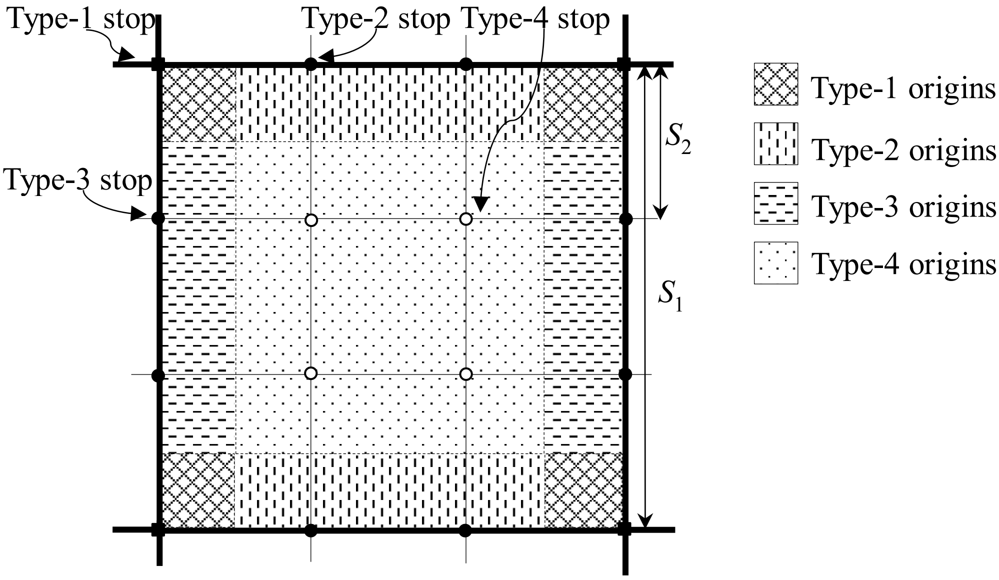


Figure 5. Probabilities of Type-*i* origins

Step 3: For each route type $\in R_k$ ($k = 1, 2, \ldots, 5$), calculate the expected trip cost for all the patrons in G_k who select I as the shortest route; i.e., $E\left[t_I | t_I = \min_{J \in R_k}(t_J)\right]$ for the OD pairs that belong to G_k. Examination of equations (3-11) shows that we only need to calculate the expected values of $X, Y, min(X, Y)$, and $max(X, Y)$ for all the ODs in G_k that select $I \in R_k$ as the shortest route. These expected values are also derived in Appendix C.

Given the above probabilities and expected values, the average patron's trip time is given by the following equation:

$$T = \sum_{k=1}^{5} \sum_{I \in R_k} P_{G_k} Q_I^{G_k} E\left[t_I |\, t_I = \min_{J \in R_k}(t_J) \text{ for OD pairs that belong to } G_k\right] \quad (15)$$

2.4 Agency cost

Here we consider four agency cost components (Daganzo, 2010a; Sivakumaran et al. 2014; Gu et al., 2016): the amortized line infrastructure cost, the amortized stop cost, the operating cost related to vehicle kilometers traveled (e.g., fuel cost), and the operating cost related to vehicle hours traveled (e.g., amortized vehicle purchase cost and staff wages). The agency cost per trip is formulated as follows:

$$AC = \frac{1}{\mu\lambda}\Big((\pi_{lo}^{L}L_{lo} + \pi_{lo}^{N}N_{lo} + \pi_{lo}^{K}K_{lo} + \pi_{lo}^{U}U_{lo}) + (\pi_{ex}^{L}L_{ex} + \pi_{ex}^{N}N_{ex} + \pi_{ex}^{K}K_{ex} + \pi_{ex}^{U}U_{ex})\Big) \quad (16)$$

where μ denotes the patron's value of time, which is used to convert the monetary cost to the temporal one (note here we assume all the patrons have the same value of time); π_{lo}^{L} and π_{ex}^{L} are the unit costs per kilometer of line infrastructure for local and express services, respectively, amortized to each hour of operation; L_{lo} and L_{ex} the local and express infrastructure lengths (for both travel directions of each line) per km^2 of service area; π_{lo}^{N} and π_{ex}^{N} the amortized unit costs per local and express stop; N_{lo} and N_{ex} the local and express stop densities per km^2; π_{lo}^{K} and π_{ex}^{K} the unit costs per vehicle-km traveled; K_{lo} and K_{ex} the vehicle-kms traveled per hour per km^2 of service area; π_{lo}^{U} and π_{ex}^{U} the unit costs per vehicle-hours traveled; and U_{lo} and U_{ex} the vehicle-hours traveled per hour per km^2 of service area. The $L_{lo}, L_{ex}, N_{lo}, N_{ex}, K_{lo}, K_{ex}, U_{lo}$, and U_{ex} are derived as follows:

$$L_{lo} = \frac{4}{S_2} - \frac{4}{S_1},\ L_{ex} = \frac{4}{S_1} \quad (17)$$

$$N_{lo} = \left(\frac{2}{S_2S_3} - \frac{1}{S_2^2}\right) - \frac{2}{S_1S_3} + \frac{2}{S_1S_2} - \frac{1}{S_1^2},\ N_{ex} = \left(\frac{2}{S_1S_2} - \frac{1}{S_1^2}\right) \quad (18)$$

$$K_{lo} = \frac{L_{lo}}{H_{lo}},\ K_{ex} = \frac{L_{ex}}{H_{ex}} \quad (19)$$

$$U_{lo} = \frac{K_{lo}}{V_{lo}},\ U_{ex} = \frac{K_{ex}}{V_{ex}} \quad (20)$$

2.5 Critical vehicle occupancies

We find that the passenger occupancy of an express vehicle is approximately invariant when the vehicle travels along the line; see Appendix D.1 for a sketched proof. Hence its maximum occupancy, O_{ex}, is equal to the total patron-kms traveled divided by the total vehicle-kms traveled, as formulated below:

$$O_{ex} = \frac{H_{ex}S_1}{4}\lambda \sum_I \sum_{k=1\ldots5} P_{G_k} Q_I^{G_k} E_{I,ex}^{G_k},\ I \in \bigcup_{k=1}^{5} R_k \quad (21)$$

where $E_{I,ex}^{G_k}$ ($I \in R_k, k = 1,2,\ldots,5$) is the expected in-vehicle travel distance on the express lines for a patron in G_k who chooses route type I. This number is calculated using the method described in Appendix C.

However, a local vehicle's passenger occupancy varies as it travels along a local line. We find that the critical occupancy of a local vehicle, O_{lo}, can be formulated as follows:

$$O_{lo} = \frac{H_{lo}S_1S_2}{4(S_1 - S_2)}\lambda\Big(P_{G_2}Q_{le}^{G_2}E_{le,lo}^{G_2} + P_{G_3}Q_{le}^{G_3}E_{le,lo}^{G_3} + P_{G_4}Q_{le}^{G_4}E_{le,lo}^{G_4} + \rho_o P_{G_2}Q_{lee}^{G_2}E_{lee,lo}^{G_2} + 2P_{G_3}Q_{lee}^{G_3}E_{lee,lo}^{G_3} + P_{G_5}Q_{ll}^{G_5}E_{ll,lo}^{G_5} + P_{G_5}Q_{lel}^{G_5}E_{lel,lo}^{G_5} + \frac{\rho_o + 2}{2}P_{G_5}Q_{leel}^{G_5}E_{leel,lo}^{G_5}\Big) \tag{22}$$

where $E_{I,lo}^{G_k}$ $(I \in R_k, k = 1,2,\ldots,5)$ is the expected in-vehicle travel distance on the local lines for a patron in G_k who chooses route type I. The coefficient ρ_o is a function of m, as presented below:

$$\rho_o = \begin{cases} \frac{3(2m-3)}{2(m-1)^3}(m^2 - 2m + 2), m \in \{2,4,6\ldots\} \\ \frac{3(m-2)}{(m-1)}, m \in \{3,5,7\ldots\} \end{cases} \tag{23}$$

The derivation of (22-23) is furnished in Appendix D.2. The $E_{I,lo}^{G_k}$ and $E_{I,ex}^{G_k}$ $(I \in R_k, k = 1,2,\ldots,5)$ can also be calculated using the method in Appendix C. Note that they apply to the cases where $m > 1$; if $m = 1$, no local line exists, and $O_{lo} = 0$.

3. Solution Method

The optimization model presented in Section 2 is a bi-level program, where the upper level is a mixed-integer nonlinear program (MINLP) furnished by (1a-h) and the lower level is the patrons' route assignment problem as described in Section 2.3. The entire problem is thus solved in an iterative manner. Specifically, the upper level problem is solved by a gradient-based search method (in this paper we use the sequential quadratic programming method implemented by the *fmincon* tool of Matlab R2016b). The lower level problem is solved by the method of successive averages (MSA; see Sheffi, 1985). To further improve the computational efficiency, we first solve the program in which the integer constraints (1f-h) are relaxed. We then search for the lowest-cost solution that satisfies (1f-h) in the neighborhood of the relaxed program's optimal solution. The detailed algorithm is furnished below:

Step 0: Randomly set the initial values of the decision variables: $S_1^{(0)}, S_2^{(0)}, S_3^{(0)}, H_{lo}^{(0)}, H_{ex}^{(0)}$. Set the outer-loop iteration count $n = 0$.

Step 1 (lower-level equilibrium): Set the inner-loop iteration count $n' = 1$. Compute the route assignment equilibrium using MSA under the current design specified by $S_1^{(n)}, S_2^{(n)}, S_3^{(n)}, H_{lo}^{(n)}, H_{ex}^{(n)}$.

Step 1.1: Calculate the trip OD probabilities $P_{G_k}^{(n')}$ and route choice probabilities $Q_I^{G_k(n')}$ $(\forall I \in R_k, k = 1, 2, \dots, 5)$ by the methods presented in Section 2.3.3 and Appendix C.

Step 1.2: Substitute $P_{G_k}^{(n')}, Q_I^{G_k(n')}$ into the formulas of t_b^{lo}, t_b^{ex} in Appendix B and compute the passenger boarding time delays per stop, $\tilde{t}_b^{lo}$, $\tilde{t}_b^{ex}$.

Step 1.3: Update the route assignment by $t_b^{lo(n')} = t_b^{lo(n'-1)} + \frac{\left(\tilde{t}_b^{lo} - t_b^{lo(n'-1)}\right)}{n'}$, $t_b^{ex(n')} = t_b^{ex(n'-1)} + \frac{\left(\tilde{t}_b^{ex} - t_b^{ex(n'-1)}\right)}{n'}$. (The $t_b^{lo(0)}$ and $t_b^{ex(0)}$ are set to zeros.)

Step 1.4: If the relative change of the equilibrium is less than a predefined tolerance ε (e.g., $\varepsilon = 0.001$), i.e., $\left|\frac{t_b^{lo(n')} - t_b^{lo(n'-1)}}{t_b^{lo(n'-1)}}\right| + \left|\frac{t_b^{ex(n')} - t_b^{ex(n'-1)}}{t_b^{ex(n'-1)}}\right| \leq \varepsilon$, we consider the solution has converged to the equilibrium and go to Step 2. Otherwise, let $n' = n' + 1$ and return to step 1.1.

Step 2 (upper-level optimization): With the current route assignment solution, and ignoring the integer constraints (1f-h), find the optimal solution to the upper level problem via the sequential quadratic programming method. Since the problem is non-convex, we repeat the optimization process with different initial solutions to ensure a good, near-optimal solution is attained. (In each numerical case that we examined, the different initial values always led to the same solution; hence we reckon that it is the optimal solution of the problem.) Set $n = n + 1$ and record the updated solution as $S_1^{(n)}, S_2^{(n)}, S_3^{(n)}, H_{lo}^{(n)}, H_{ex}^{(n)}$. If the convergence criterion of the search method is satisfied, report the current solution: $\tilde{S}_1 = S_1^{(n)}, \tilde{S}_2 = S_2^{(n)}, \tilde{S}_3 = S_3^{(n)}, \widetilde{H}_{lo} = H_{lo}^{(n)}, \widetilde{H}_{ex} = H_{ex}^{(n)}$, and the optimal objective function value $\tilde{Z} = Z^{(n)}$. Otherwise, return to Step 1.

Step 3: For each of the four possible combinations of m and m', which take the floor and ceiling of $\frac{\tilde{S}_1}{\tilde{S}_2}$ and $\frac{\tilde{S}_2}{\tilde{S}_1}$, respectively, rerun Steps 1-2 to obtain an updated solution that satisfies (1f-1g). Record the lowest-cost solution among them.

Step 4: Repeat Step 0-3 for many times with different initial values of the decision variables, until the lowest-cost solution stays invariant for a number of times. (For each numerical instance that we examined, we repeated Step 0-3 by 10 times and the 10 solutions were always the same. Thus, we believe the global optima have been attained.)

4. Numerical Case Studies

In this section, we examine the optimal designs of the intersecting bimodal network for a variety of generic cities in the world. Three typical transit modes (ordinary bus, BRT, and rail) are considered. The parameter values are furnished in Section 4.1. Sections 4.2-4.4 present three batteries of comparisons between the optimal intersecting bimodal network and, respectively, i) the optimal single-mode grid network (Holroyd, 1967; Daganzo, 2010b); ii) the trunk-feeder network proposed by Sivakumaran et al. (2014); and iii) a combination of express and local networks that are separately optimized. The last battery of comparisons unveils the necessity of jointly optimizing the two transit modes.

4.1 Parameter values

We consider the demand density λ spanning over a wide range of $[100, 1000]$ trips/km^2/h, the average trip length $L \in [5, 20]$ km, and the value of time μ taking two values: 5 \$/h for a low-wage city (e.g. Guangzhou, China), and 20 \$/h for a high-wage city (e.g. Barcelona, Spain). Table 2 summarizes the operational and cost parameter values for the three transit modes, which are borrowed from the previous studies (Vuchic, 2007; Daganzo, 2010a; Sivakumaran et al., 2014; Gu et al., 2016). In addition, we specify the intra-modal transfer penalties as $[\xi_{ll}^{\text{bus}}, \xi_{ee}^{\text{BRT}}, \xi_{ee}^{\text{Rail}}] = [30,40,60]$ seconds (Daganzo, 2010a; Sivakumaran et al., 2014), and the inter-modal transfer penalties as $[\xi_{le}^{\text{bus-BRT}}, \xi_{le}^{\text{bus-Rail}}, \xi_{le}^{\text{BRT-Rail}}] = [60,90,90]$ seconds. The m_0 in constraint (1f) is set to be 4.[3]

[3] This value is set to ensure the assumption $L \gg S_1$ is true, so that the proofs in Appendix D are valid, and the effect of backtracking (which may occur for route types eee and lel) is small. For example, consider an OD type (2,2) where the origin and destination are both located between the same two neighboring N-S express lines. If route type eee is selected, then the trip involves backtracking and the associated cost is underestimated by our present models. But when $m_0 = 4$, we find that the percentage of trips involving this type of backtracking is below 3% for all the numerical instances that we examined. Thus, the modest underestimation does not compromise the validity of our cost models and findings.

Table 2. Parameter values for the numerical case studies

Parameters \ Transit mode	Bus	BRT	Rail
V_w, km/h	2	2	2
v_{lo}, v_{ex}, km/h	25	40	60
t_d^{lo}, t_d^{ex}, h/stop	0.0083	0.0083	0.0125
τ, s/patron	2	1	0
C_{lo}, C_{ex}, patron/vehicle	80	150	2400
$H_{lo}^{min}, H_{ex}^{min}$, h	0.05	0.033	0.033
π^L, \$/km	$6 + 0.2\mu$	$162 + 5.4\mu$	$594 + 19.8\mu$
π^N, \$/stop	$0.42 + 0.014\mu$	$4.2 + 0.14\mu$	$294 + 9.8\mu$
π^K, \$/vehicle-km	0.59	0.66	2.20
π^U, \$/vehicle-h	$2.66 + 3\mu$	$3.81 + 4\mu$	$101 + 5\mu$

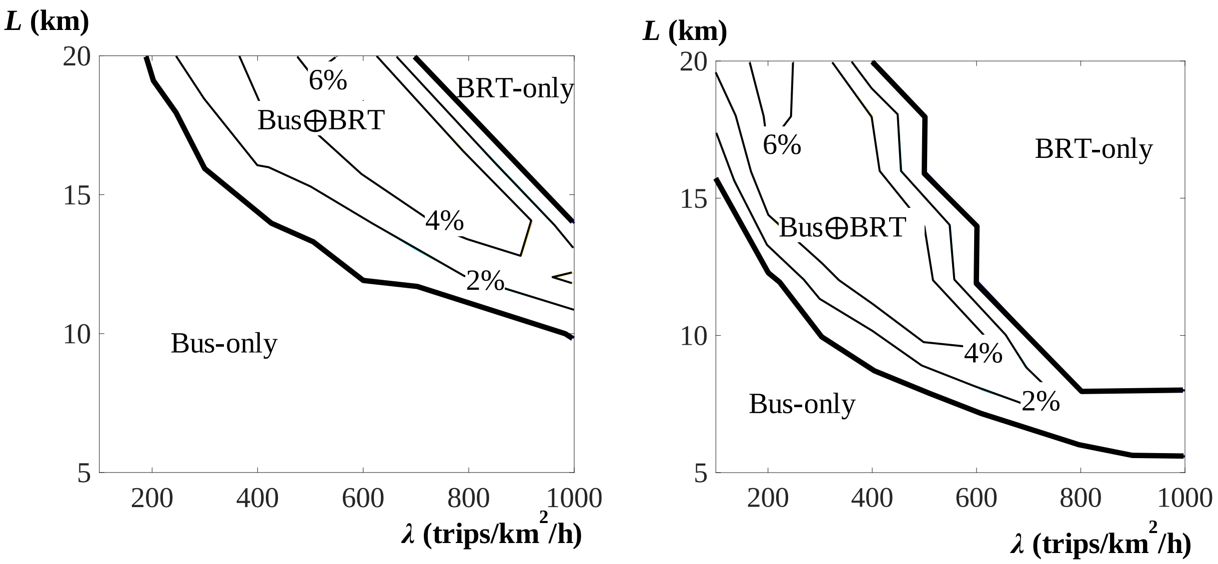
4.2 Optimal designs of intersecting bimodal networks

We denote $\mathcal{M}_L \oplus \mathcal{M}_E$ as the intersecting bimodal network where the local mode is $\mathcal{M}_L$ and the express mode is $\mathcal{M}_E$, $\mathcal{M}_L$ and $\mathcal{M}_E \in \{\text{Bus}, \text{BRT}, \text{Rail}\}$. We hereby report the results for two types of intersecting networks: Bus⊕BRT and Bus⊕Rail. The other combinations of $\mathcal{M}_L$ and $\mathcal{M}_E$ are either inferior to the above two, or reducing to single-mode networks (i.e., $m = 1$) at the optimality. The two intersecting networks are compared against three single-mode networks (Bus-only, BRT-only, and Rail-only). For each single-mode network, we assume that the transit lines form a grid network with a line spacing that is an integer multiple of the stop spacing (Holroyd, 1967; Daganzo, 2010b).

We first look at generic BRT-cities where only BRT and ordinary bus are considered, while rail is not an option. Figures 6a and b show the lowest-cost design among the Bus⊕BRT, BRT-only, and Bus-only networks over the range of λ and L defined in Section 4.1 for a low-wage and a high-wage city, respectively. The regions where each network type wins are divided by the thick solid curves. Both figures show that the intersecting bimodal design outperforms the single-mode networks where λ and L are moderately high. When λ and L are both very high, the optimal bimodal network reduces to a BRT-only network (i.e. $m = 1$). The Bus-only network turns out to triumph where λ or L is low.

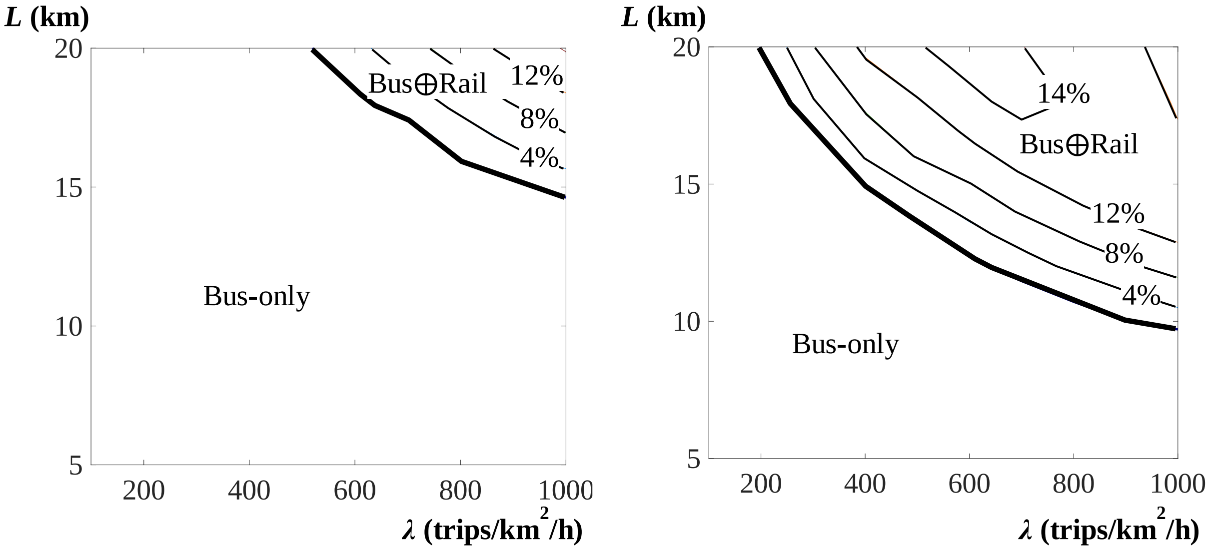
The relative generalized cost savings of the intersecting bimodal design, as defined by $\left(1-\frac{Z_{\mathrm{Bus}\oplus\mathrm{BRT}}}{\min\{Z_{\mathrm{Bus-only}},Z_{\mathrm{BRT-only}}\}}\right)\times 100\%$, are plotted by the thin contour lines in the regions dominated by Bus⊕BRT in Figures 6a and b. These contours show cost savings of up to 6% in both figures. The savings in the patrons' travel cost (T) are even higher: they are up to 11-13% (not shown in the figures). This is as expected, because an intersecting bimodal network has redundancy (i.e. offering multiple route options to the patrons), which benefits the patrons. It also implies that our predictions of the benefits of the intersecting networks are conservative, because lower travel costs will induce demand shift from other modes to transit, and this is not taken into account in our models.

For generic rail-cities where the Bus⊕Rail, Bus-only, and Rail-only networks are considered, smaller regions of dominance are observed for the Bus⊕Rail network; see Figures 7a and b for the low- and high-wage cities, respectively. The generalized cost savings, however, are even higher (up to 14%), as shown by the thin contour lines in the figures. Compared to Figures 6a and b, now the Bus-only network wins in larger regions of λ and L, while the Rail-only network attains the lowest cost only when λ is higher than the maximum demand density examined here (thus it cannot be seen in the figures).



a. Low-wage city ($\mu = 5\$/h$) b. High-wage city ($\mu = 20\$/h$)

Figure 6. Lowest-cost design between Bus⊕BRT, Bus-only, and BRT-only networks



a. Low-wage city ($\mu = 5\$/h$)   b. High-wage city ($\mu = 20\$/h$)

Figure 7. Lowest-cost design between Bus⊕Rail, Bus-only, and Rail-only networks

One may be also interested in the optimal ratio between the sizes (infrastructure lengths) of the local and express networks (i.e. the value of m). Figures 8a and b plot the contour lines of the optimal m for the Bus⊕BRT network for the low- and high-wage cities, respectively; and Figures 9a and b plot the same contours for the Bus⊕Rail network. For most scenarios, the optimal m ranges from 1 to 4, i.e., the optimal bimodal design will render the express line spacing being two to four times of that of the local lines. The figures also reveal that as λ or L increases, the optimal m generally decreases; i.e., the proportion of the express network grows with the demand, which is as expected. An exception occurs in Figure 9a, where the optimal m increases with L. This exception is caused by the integer constraint (1f).

The results of patrons' route choices exhibit high variability and diversity. For example, for an optimal Bus⊕BRT network in the low-wage city (with the demand in its dominance region as shown in Figure 6a), the proportion of patrons who take an express-only route (of types ee and eee) varies between 12% and 55%. The proportion of those who take an intermodal route (of types le, lee, lel, and $leel$) ranges from 45% to 87%. The lowest proportion of intermodal routes (45%) and the highest proportion of express-only routes (55%) occur when the λ and L are both the highest (i.e., when $m = 2$ and the express lines contribute to half of the bimodal network); and the highest proportion of intermodal routes (87%) and the lowest proportion of express-only routes (12%) occur when the λ is low (i.e., when m is the largest). The high proportions of intermodal routes manifest that both modes are essential in the bimodal system. In addition, 60-86% of the local trip segments are feeder trips, and 14-40% are for direct travel (in either E-W or N-S direction). This confirms that the two roles served by the local lines, i.e. as feeder service for long trips and as direct service for shorter trips, are both significant. On the other hand, the proportion of patrons who take local lines only (route type ll) is generally below 1%.

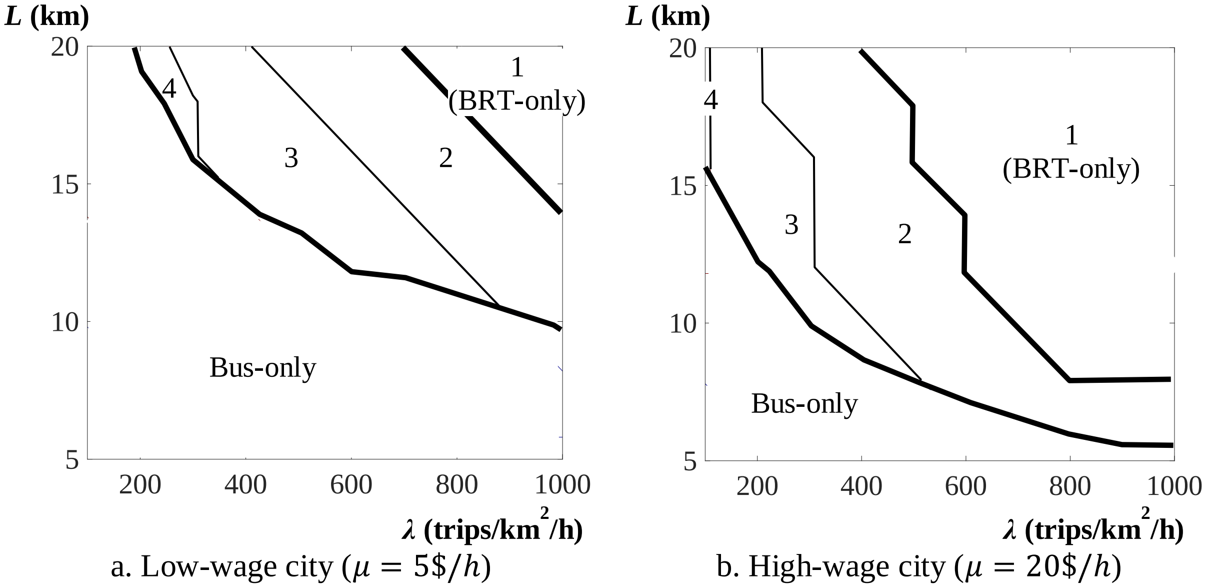


Figure 8. Optimal m of Bus⊕BRT network

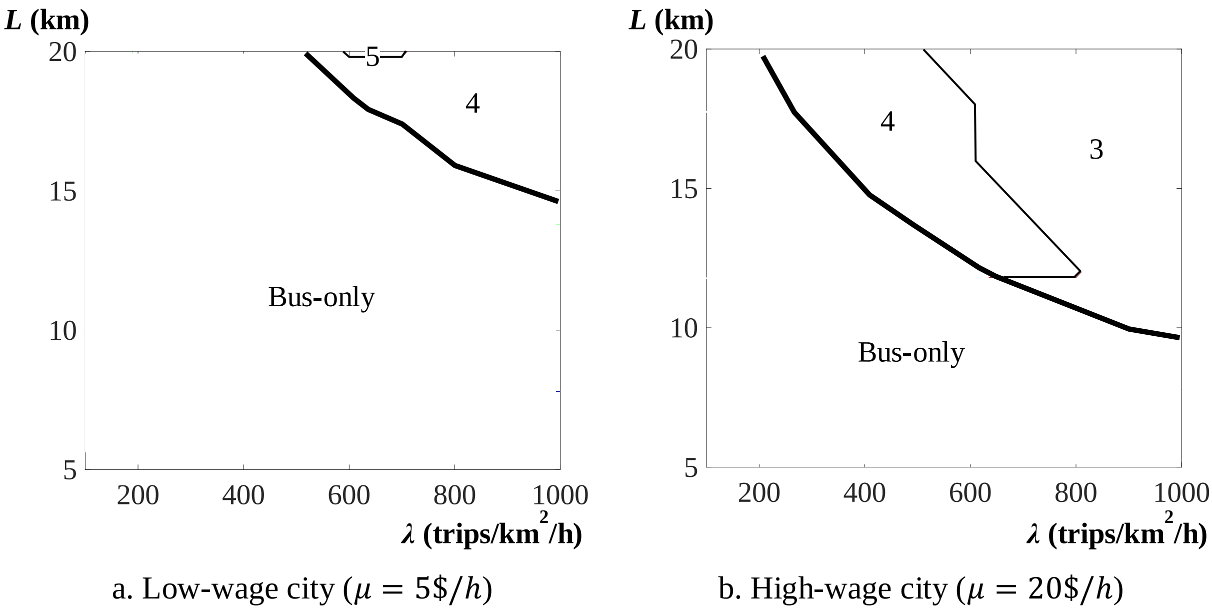
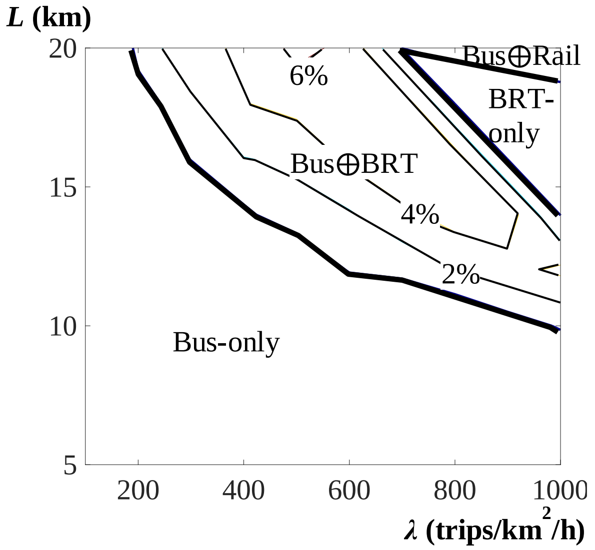


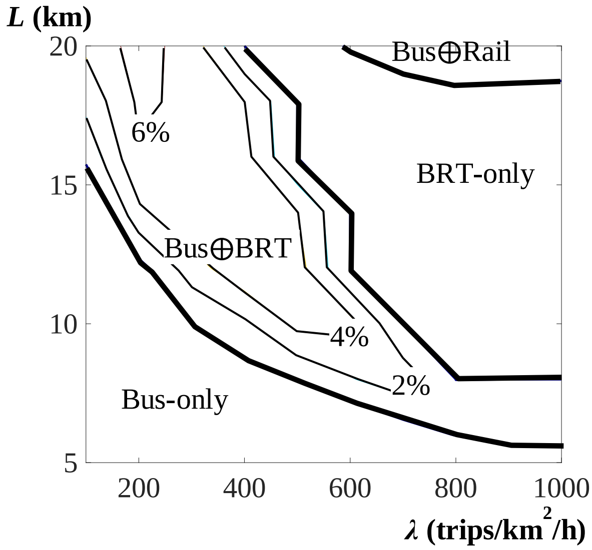
Figure 9. Optimal m of Bus⊕Rail network

We next consider a scenario where all the five network types (Bus⊕BRT, Bus⊕Rail, Bus-only, BRT-only, and Rail-only) are included in the menu for comparison (i.e. the transit planner does not specify the express mode). The lowest-cost designs are plotted in Figures 10a and b for low- and high-wage cities, respectively. The generalized cost savings of the optimal intersecting network, defined by $\left(1-\frac{\min\{Z_{\mathrm{Bus}\oplus\mathrm{BRT}},\ Z_{\mathrm{Bus}\oplus\mathrm{Rail}}\}}{\min\{Z_{\mathrm{Bus-only}},\ Z_{\mathrm{BRT-only}},\ Z_{\mathrm{Rail-only}}\}}\right)\times 100\%$, are again plotted by the thin contour lines. The figures are similar to Figures 6a and b, except that now Bus⊕Rail appears to be the optimal design for very high values of λ and L. This reveals the great competitiveness of the Bus⊕BRT network (and the BRT-only network as a special case of the Bus⊕BRT network).

It is not surprising to see that the bimodal designs outperform single-mode designs under certain conditions since the two transit modes can complement each other to some degree. It would be more interesting to see how the intersecting bimodal network performs as compared to other bimodal designs. We next compare the optimal intersecting design against the trunk-feeder network, which was proposed by Sivakumaran et al. (2014).



a. Low-wage city ($\mu = 5\$/h$)



b. High-wage city ($\mu = 20\$/h$)

Figure 10. The optimal design from scratch

4.3 Optimal intersecting network versus trunk-feeder network

The trunk-feeder network is illustrated in Figure 11, where the trunk lines (BRT or rail) form a grid structure, and the feeder lines (bus) transport the patrons to and from the trunk line stations. In this network, every patron has to take a feeder bus to access the nearest trunk line station, travel in the trunk network, and eventually transfer to another feeder bus to reach her destination. Traveling by trunk lines only is not a feasible option for all but a very small portion of patrons because the access cost is high. Traveling by feeder buses only is also undesirable (even in one of the E-W and N-S directions only) since the feeder buses take detours. The design variables to be optimized here include the line spacings and headways of trunk and feeder lines, and the feeder stop spacing.

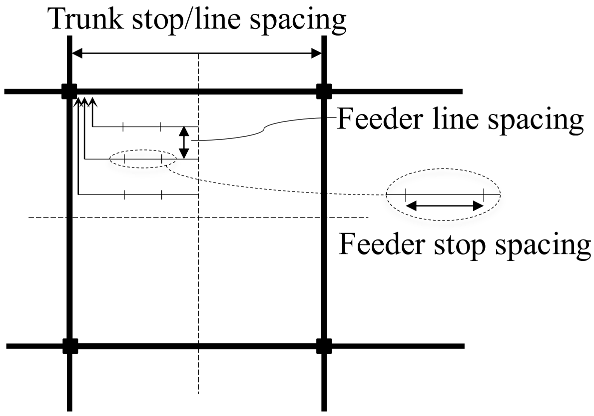
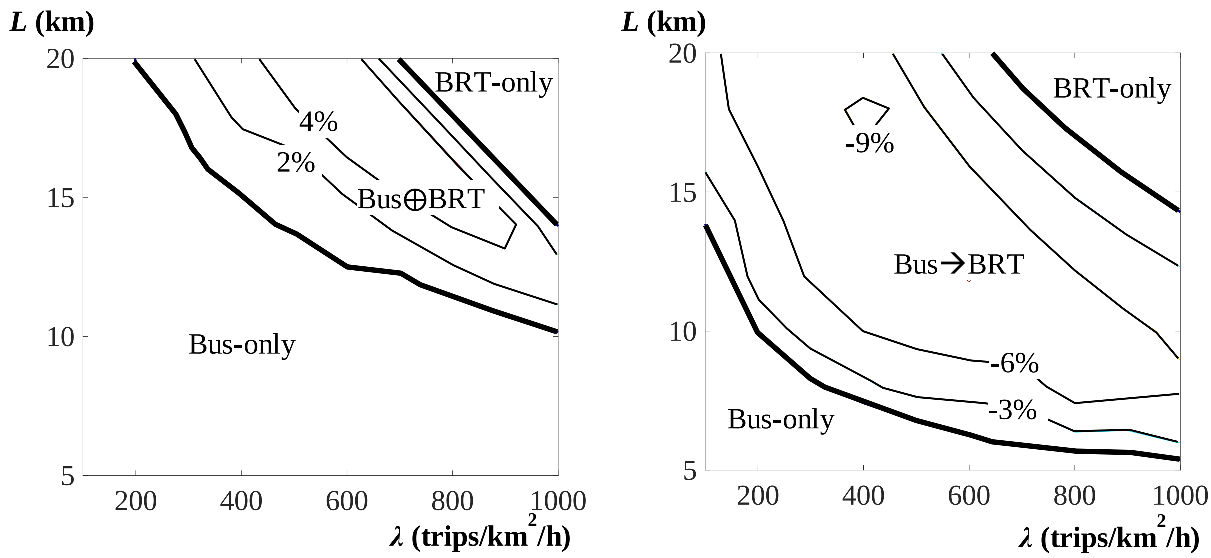


Figure 11. A trunk-feeder network (Sivakumaran et al., 2014)

For illustration, we plot the optimal network type for the same range of λ and L in Figures 12a and b for low- and high-wage cities, respectively, where BRT is selected to be the express mode for the intersecting network and the trunk mode for the trunk-feeder network. The contour lines show the percentage of generalized cost savings by Bus⊕BRT as compared against the trunk-feeder network (denoted by Bus→BRT), i.e. $\left(1 - \frac{Z_{\text{Bus}\oplus\text{BRT}}}{Z_{\text{Bus}\rightarrow\text{BRT}}}\right) \times 100\%$. We find that the Bus⊕BRT network wins in the low-wage city, while the Bus→BRT network triumphs in the high-wage city. This is because the trunk-feeder network has a lower patron cost but a higher agency cost, and the latter has a lower weight in the generalized cost function in a high-wage city. We also find that if rail is selected as the express and trunk mode, the intersecting network Bus⊕Rail will completely lose out to the trunk-feeder network. This is not surprising too, because in a bimodal network with bus and rail, the slow and cheap buses are more suitable for serving as feeders. (Comparison between the optimal solutions of Bus⊕BRT and Bus⊕Rail also unveils that the proportion of local feeder trips is much higher for the latter design.)



a. Low-wage city ($\mu = 5\$/h$) b. High-wage city ($\mu = 20\$/h$)

Figure 12. Lowest-cost design between Bus⊕BRT, Bus→BRT, Bus-only, and BRT-only networks

The trunk-feeder design often performs better than the intersecting design in terms of the generalized cost because of the following two reasons:

(i) The transfer penalties are set to be low in our numerical cases, while in reality the penalties could be much higher. For example, Guo and Wilson (2004) reported that an average transit patron's perceived penalty per transfer (including additional walking time, unreliability of the connecting service, risk of losing a seat, etc.) is equivalent to 5-20 minutes of in-vehicle travel time (see Tables 1 and 8 in the cited work). Moreover, each additional transfer in a trip would impose a higher penalty to the patron (Liu et al. 1997). Hence, the low penalty values used in this paper is in favor of the trunk-feeder network, because the trunk-feeder network has 3 transfers per trip, while the intersecting network has less than 2 transfers per trip. On the other hand, the intersecting network can be more advantageous over the trunk-feeder network if higher transfer penalty values are used. For example, Figure 13 illustrates the lowest-cost designs for a high-wage city with $[\xi_{ll}^{\text{bus}}, \xi_{ee}^{\text{BRT}}, \xi_{ee}^{\text{Rail}}] = [420,120,120]$ seconds and $[\xi_{le}^{\text{bus-BRT}}, \xi_{le}^{\text{bus-Rail}}, \xi_{le}^{\text{BRT-Rail}}] = [180,180,180]$ seconds, which are aligned with the Stockholm case reported by Guo and Wilson (2004). By comparing against Figure 12b, note in Figure 13 that Bus⊕BRT outperforms Bus→BRT for cases with low trip lengths.

(ii) The intersecting network provides multiple route options. Redundant network designs like this are usually not cost-effective. To the contrary, the trunk-feeder network specifies only one route option (feeder-trunk-trunk-feeder) for all the patrons. Consequently, the patron flows are more concentrated on the trunk lines of a trunk-feeder network than on the express lines of an intersecting network, given the same demand level. (This is manifested by our numerical results.) Hence the trunk-feeder design is naturally more cost-effective due to the economy of demand concentration for transit systems (Chen et al., 2015). However, a redundant network also has merits, e.g. greater resilience when part of the network is heavily congested or failed. This is because the affected patrons can easily find convenient alternative routes in a partially failed transit network. Resilience is often a major concern in transit network planning, although we intend not to furnish a full analysis of network resilience here.

In light of the above, the numerical results presented here are conservative to the intersecting network, and this design should exhibit more advantages in reality as compared against the trunk-feeder network.

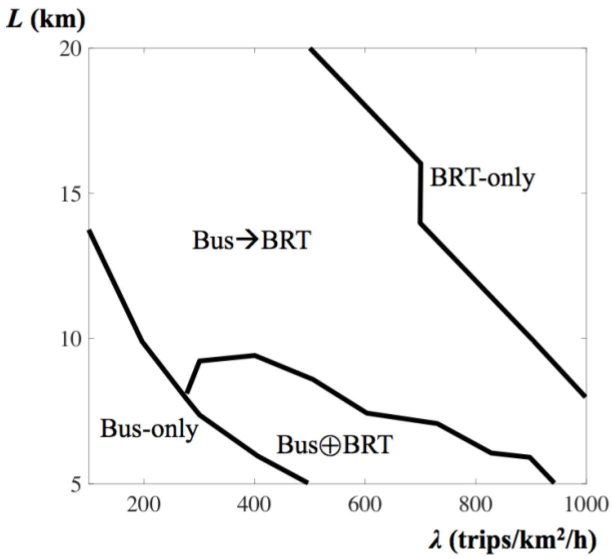


Figure 13. Lowest-cost design with higher transfer penalties in a high-wage city ($\mu = 20\$/h$)

4.4 Joint design versus separate design of bimodal networks

In this section, we show that the advantages of the proposed intersecting bimodal networks can be achieved only by jointly optimizing the two transit modes.

In practice, two major transit modes of a city (e.g. rail and bus) are often planned and designed separately and at different times. Planning of a new mode (e.g. rail) is generally based upon a fixed demand (i.e. OD) forecasted using survey data, and thus fails to properly model how the existing and new patrons choose modes and routes according to the new bimodal network layout. Albeit some minor adjustments are applied to the existing mode, e.g., removal or relocation of some existing bus lines that were overlapping with a new metro line, these adjustments are decided according to rules of thumb, instead of being optimized systematically. To see how this conventional "separate design" process undermines the benefits of bimodal networks, we compare the total cost of separately designed bimodal networks against the optimal cost of the intersecting network. To be conservative, the cost of the separately designed networks for serving a combined demand of density λ is calculated as follows:

(i) For an arbitrary split ratio $\alpha \in [0.1, 0.9]$, we design two optimal single-mode grid networks (one for the local mode and the other for the express mode) to serve the split demand of densities $\alpha\lambda$ and $(1-\alpha)\lambda$, respectively.

(ii) In the optimal designs of the two separate networks, we remove the local lines that overlap with any express line, and recalculate the agency costs. We also recalculate the average patron's cost by employing the route choice models presented in Section 2.3.[4]

(iii) We find the optimal value of α that minimizes the generalized cost per trip for the two single-mode networks combined. This lowest cost is recorded as the generalized cost for the separate design.

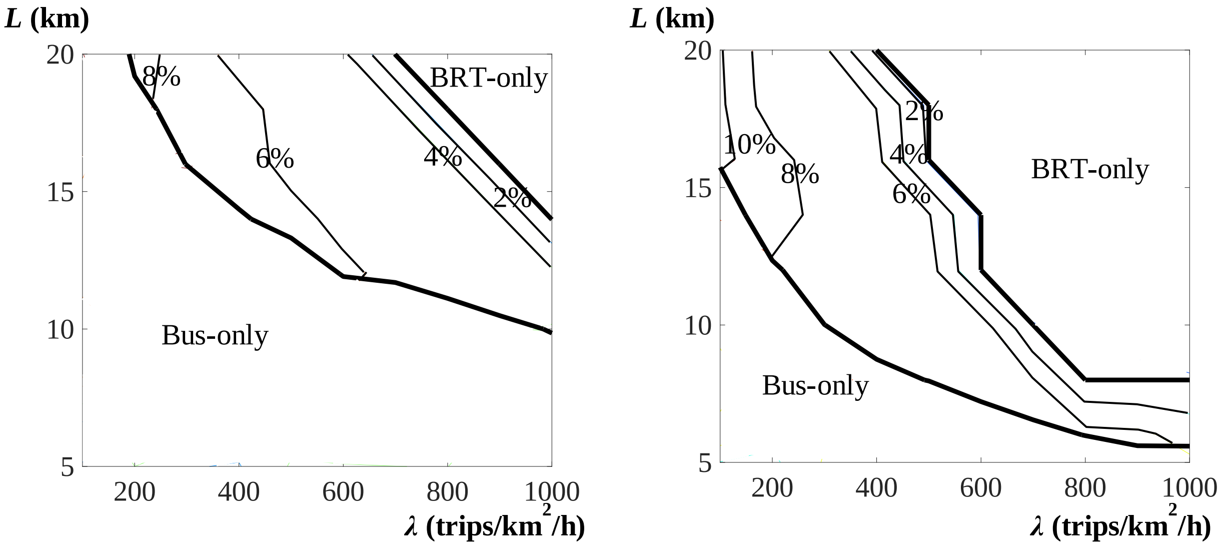
Figures 14a and b demonstrate, for the Bus⊕BRT network in low- and high-wage cities respectively, the contours of percentage of generalized cost increase by optimizing the two modes separately, i.e., $\left(\frac{Z^{\text{separate}}}{Z^{\text{joint}}} - 1\right) \times 100\%$, where Z^{separate} and Z^{joint} are the generalized costs for the separate designs and the joint intersecting network design, respectively. These contour lines show that the separately designed bimodal networks, albeit being optimized under a number of favorable assumptions described above, increase the

[4] Since the two single-mode networks are separately optimized, the local lines usually do not overlap with any express line in the bimodal network. However, to make a conservative comparison, we "pretend" that every express line "overlaps" with a local line, and remove the cost of those "overlapping" local lines from the agency cost. We still keep the local line spacing unchanged when calculating the patrons' costs, and ignore the extra walking distance when a patron transfers between an express and a local stations. Note that the two stations usually do not overlap in the separately-designed network. All this results in underestimation of the generalized cost for the separate design.

generalized system cost by 2-10% as compared against the optimal joint design for both low- and high-wage cities. For the Bus⊕Rail network, the cost increases by the separate designs are even higher (not shown). Hence, the benefit of the bimodal network (see Figures 6a and b) almost vanishes if the two modes are separately designed. This implies that the present bimodal transit systems in many cities may be cost-ineffective, and significant cost savings can be achieved by simply implementing a joint design process.

5. Conclusions

This paper proposes continuum models for minimizing the generalized cost of a bimodal transit system that consists of two intersecting unimodal networks. To our best knowledge, this is the first work to study a general form of bimodal transit networks with a realistic level of line redundancy, although similar transit systems are commonly found in the real world. Unlike other transit networks studied in the literature, the intersecting bimodal network furnishes multiple route options for the patrons to choose, and this poses new challenges to the modeling effort. To address the challenge, we formulate the problem as a bi-level program, where the patrons' mode and route assignment problem is solved at the lower level. We are also the first to model transit patrons' realistic route choice in a multimodal context by combining continuum models for transit network design with user equilibrium solutions. The work unveils new insights for both practice and theoretical development, which are summarized below.



a. Low-wage city ($\mu = 5\$/h$) b. High-wage city ($\mu = 20\$/h$)

Figure 14. Percentages of generalized cost increases by designing the intersecting Bus and BRT networks separately

First of all, our numerical results show that the intersecting network outperforms the unimodal networks for medium to high demand levels (see Figures 6 and 7), and the trunk-feeder network for low-wage cities when BRT is the express/trunk mode (see Figure 12a). This manifests that the network with redundancy can also be cost-effective under certain operating conditions. The benefit in real cities could be even larger, since the network redundancy favors the transit patrons, which will thus induce more travelers to take the transit. On the other hand, the local-only and express-only unimodal networks furnish the lowest-cost designs under low and very high demand, respectively; see again Figures 6 and 7. This means bimodal networks (including the trunk-feeder design) are not suitable for cities under these demand levels. For rich cities and rail-cities, the trunk-feeder structure outperforms the intersecting network structure if the transfers are made convenient. Yet the intersecting network can perform better in real practice where the transfer penalty is higher, and where network resiliency is taken into account.

Also note how important it is to jointly optimize the design of a bimodal transit network, since the conventional separate design procedure can significantly undermine the benefit of bimodal networks; see Figures 14a and b. This implies that when a city plans its second transit mode (e.g. rail), the existing mode (e.g. bus)'s network layout, stop locations and service frequencies should be adjusted according to the joint optimal design of the two modes.

Methodologically, our work has extended the theory of continuum models for transit network optimization by accounting for more realistic route choices of the patrons within the bimodal transit network. Additional numerical tests, whose details are not reported in this paper for brevity, have confirmed the necessity of integrating the patrons' route choices into the modeling work, since otherwise the generalized cost would be overestimated by up to 14%. Our formulation and the bi-level approach can be further extended to solve more complicated problems in the realm of transit network design. Some examples include: the problems i) that appreciate stochastic mode and route choices of the patrons; ii) that incorporate alternative access modes (e.g. bikes and flexible-route feeder buses); and iii) that involve competitive non-transit modes (e.g. cars) and elastic demand. Select topics are currently under investigation.

Our assumptions regarding a symmetric, idealized grid network and the static, uniform demand pattern are necessary for developing the parsimonious formulation, and for examining how the optimal network design is influenced by key operating parameters that can take a variety of values. For cities that have roughly flat demand patterns, the idealized optimal design obtained from our models can be implemented with moderate local amendments of line and stop locations; see Estrada et al. (2011) for an example in Barcelona, Spain. The methodology presented in this paper can also be extended to model bi-modal transit networks for cities under spatially heterogeneous demand patterns. Work is ongoing in this regard.

Acknowledgements

The research was supported by a General Research Fund (Project No. 15280116) provided by the Research Grants Council of Hong Kong and the National Natural Science Foundation of China (Project No. 51608455). We thank Dr. Xin Li of the Hong Kong Polytechnic University for his valuable comments during the preparation of this manuscript.

Appendix A. Notation

Table A1. Notation

Notation	Description	Unit
Decision variables		
S_1	Line spacing of express system	km
S_2	Stop spacing of express system (line spacing of local system)	km
S_3	Stop spacing of local system	km
H_{lo}	Headway of local lines	hour
H_{ex}	Headway of express lines	hour
Other variables and parameters		
α	Demand split ratio for separate design of bimodal networks	
AC	The agency cost	hour/patron/hour
b_I	Number of boarding patrons per stop for route type $I \in \bigcup_{k=1,\ldots,5} R_k = \{ee, le, lee, eee, ll, lel, leel\}$	patron/stop
B_{lo}, B_{ex}	Number of boarding patrons per local and express stop	patron/stop
$d_C^{G_k}$	Critical distance at which a patron in OD type G_k is indifferent between two route options	km
D_I	Transit demand of route type $I \in \bigcup_{k=1,\ldots,5} R_k$	trip/hour/km^2
ξ_J	Transfer penalty of type $J \in \{ee, ll, le/el\}$	hour
f_I^{lo}, f_I^{ex}	Number of boardings per patron in the local and express networks, respectively, for route type $I \in \bigcup_{k=1,\ldots,5} R_k$	
G_k	Group of OD types, $k = \{1, \ldots, 5\}$	
I	Type of routes, $I \in \bigcup_{k=1,\ldots,5} R_k$	
$Q_I^{G_k}$	Probability of choosing route type I for a patron in G_k	
L	Maximum length of a trip in the N-S and E-W directions	km
m, m'	Integer ratios between S_1 and S_2, and between S_2 and S_3, respectively	
O_{lo}, O_{ex}	Critical occupancies for local and express vehicles	patron/vehicle
P_{G_k}	Probability of a trip belonging to OD group G_k	
R_k	Set of route options for OD types in G_k, $k = \{1, \ldots, 5\}$	
t_b^{lo}, t_b^{ex}	Patrons' boarding time per stop for local and express vehicles	hour
t_d^{lo}, t_d^{ex}	Constant delays per stop due to acceleration and deceleration for local and express vehicles	hour
t_I	Travel time on route type $I \in \bigcup_{k=1,\ldots,5} R_k$	hour
T	The average patron's travel time (the patron's cost)	hour
v_{lo}, v_{ex}	Cruise speeds of local and express vehicles	km/hour
V_{lo}, V_{ex}	Commercial speeds of local and express vehicles	km/hour
V_w	Walking speed	km/hour
Z	Generalized cost per patron of an intersecting network	hour
λ	Demand density	trip/hour/km^2
μ	Patrons' value of time	\$/hour
τ_{lo}, τ_{ex}	Unit boarding times per patron for local and express vehicles	hour
π_{lo}^L, π_{ex}^L	Unit costs for the local and express line infrastructure	\$/km
π_{lo}^N, π_{ex}^N	Unit costs for a local and an express stop	\$/stop
π_{lo}^K, π_{ex}^K	Unit operating costs related to vehicle-kms traveled in local and express networks	\$/vehicle-km
π_{lo}^U, π_{ex}^U	Unit operating costs related to vehicle-hours traveled in local and express networks	\$/vehicle-hour
C_{lo}, C_{ex}	Patron-carrying capacities for local and express vehicles	patron/vehicle
$H_{lo}^{min}, H_{ex}^{min}$	Minimum headways for local and express vehicles	hour

Appendix B. Patrons' boarding and alighting times per stop

We assume that t_b^{lo} and t_b^{ex} are dictated by the time needed for the boarding of patrons at a stop, because the alighting process at the same stop usually takes much less time (Newell, 1994). Thus, we have:

$$t_b^{lo} = B_{lo}\tau_{lo} \tag{B1}$$

$$t_b^{ex} = B_{ex}\tau_{ex} \tag{B2}$$

where τ_{lo} and τ_{ex} are the unit boarding times per patron for local and express vehicles, respectively; and B_{lo} and B_{ex} denote the average numbers of boarding patrons per stop for local and express vehicles, respectively[5].

The B_{lo} and B_{ex} can be obtained by summing up the numbers of boarding patrons that take different types of routes:

$$B_{lo} = b_{ll}^{lo} + b_{le}^{lo} + b_{lee}^{lo} + b_{lel}^{lo} + b_{leel}^{lo} \tag{B3}$$

$$B_{ex} = b_{ee}^{ex} + b_{le}^{ex} + b_{lee}^{ex} + b_{eee}^{ex} + b_{lel}^{ex} + b_{leel}^{ex} \tag{B4}$$

where b_I^{lo} and b_I^{ex} $(I \in \{ee, le, lee, eee, ll, lel, leel\}$ are the average boarding numbers per stop for route type I for local and express vehicles, respectively.

For each I, b_I^{lo} (or b_I^{ex}) is calculated by summing up the boarding demand at both origin and transfer stops on the local (or express) lines, divided by the corresponding vehicle-kms traveled, and then multiplied by the stop spacing. Specifically, we have:

$$b_I^{lo} = \lambda f_I^{lo} \sum_{k=1}^{5} P_{G_k} Q_I^{G_k} \frac{S_3}{K_{lo}}, I \in \{ll, le, lee, lel, leel\} \tag{B5}$$

$$b_I^{ex} = \lambda f_I^{ex} \sum_{k=1}^{5} P_{G_k} Q_I^{G_k} \frac{S_2}{K_{ex}}, I \in \{ee, le, lee, eee, lel, leel\} \tag{B6}$$

where P_{G_k} $(k = 1, 2, \dots, 5)$ denotes the probability that a trip belongs to OD group G_k, and $Q_I^{G_k}$ $(I \in \bigcup_{k=1}^{5} R_k, k = 1,2,\dots,5)$ denotes the probability that a patron in G_k chooses route option I. Thus $\sum_{k=1}^{5} P_{G_k} Q_I^{G_k}$ $(I \in \bigcup_{k=1}^{5} R_k)$ is the probability that an arbitrary patron chooses route type I. The P_{G_k} and $Q_I^{G_k}$ are derived in Section 2.3.3 and Appendix C. The K_{lo} and K_{ex} are the local and express vehicle-kms traveled per hour per km^2 of service area, which are given in (19) in Section 2.4. Finally, f_I^{lo} and f_I^{ex} $(I \in \bigcup_{k=1}^{5} R_k)$ represent the numbers of boardings on

[5] For local lines, the number of boarding patrons varies across the stops (see Appendix D). However, this variation has a very small effect on the vehicles' commercial speed. Hence, here we still use the average boarding numbers per stop to calculate t_b^{lo} and t_b^{ex} for local vehicles.

the local and express lines, respectively, for a trip of route type I. They take the following values:

$$f_{le}^{lo} = f_{lee}^{lo} = 1, f_{ll}^{lo} = f_{lel}^{lo} = f_{leel}^{lo} = 2; f_{le}^{ex} = f_{lee}^{ex} = 1, f_{ee}^{ex} = f_{lee}^{ex} = f_{leel}^{ex} = 2, f_{eee}^{ex} = 3 \qquad \text{(B7)}$$

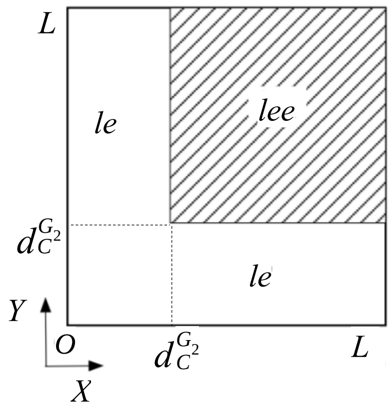
Note that if $S_2 = S_1$, K_{lo} will be zero, implying that local lines won't be built, then B_{lo} is also set as zero.

Appendix C. Route choice probabilities and expected travel distances

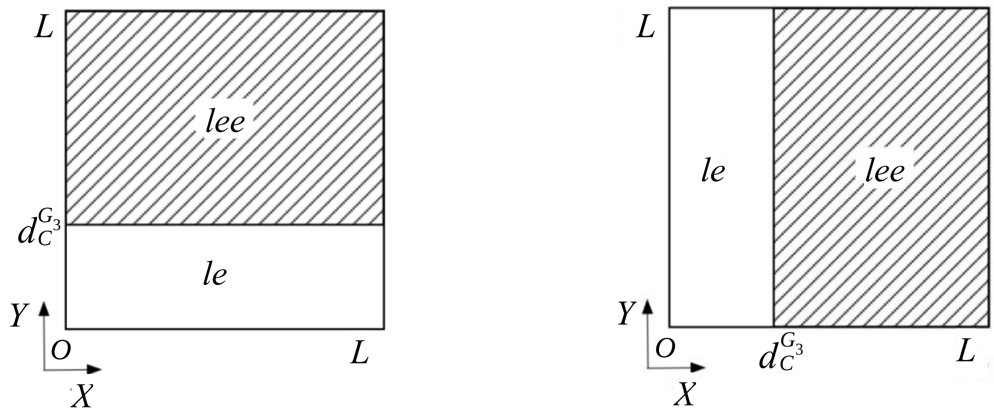
Consider an arbitrary trip origin O, the destinations of all the trips originated from O are uniformly distributed in a square of $2L$ by $2L$ centered at O. Due to the symmetry, we only need to look at the upper-left quadrant of this square. Figures C1a-d show this quadrant for each OD type that belongs to G_k ($k = 2,3,4,5$), respectively, which has more than one route options. Each square-shaped quadrant in the figures is divided into 2 or 3 regions with different shadings and route type labels; and each region labeled by route type $I \in \{le, lee, eee, ll, lel, leel\}$ is the set of trip destinations for which route type I has the shortest travel time, i.e., $t_I(X,Y) = \min_{J \in R_k}\left(t_J(X,Y)\right)$ for the OD type that belongs to $G_k, k = 2,3,4,5$. Take OD type $(1,4)$ in G_2 as an example (Figure C1a), where each patron chooses between route types le and lee. The boundaries between the destination regions for le and lee are where $t_{le} = t_{lee}$. Using the route travel time models (4-5) and (11a-b) in Section 2.3.2, we find that a patron will choose lee if $\min(X,Y) \geq d_C^{G_2} \equiv \left(\xi_{ee} + \frac{H_{ex}}{2} + \frac{S_1}{4V_{lo}}\right)/\left(\frac{1}{V_{lo}} - \frac{1}{V_{ex}}\right)$, and will choose le otherwise, as shown in the figure. The destination regions for each route option for OD types (2,4), (3,4), (2,2), and (3,3) can be similarly obtained, as illustrated in Figures C1b and c, where the critical distances $d_C^{G_3} = d_C^{G_2}$, and $d_C^{G_4} = \left(\frac{S_2 - S_3}{4V_w} + H_{ex} - \frac{H_{lo}}{2} + 2\xi_{ee} - \xi_{le}\right)/\left(\frac{1}{V_{lo}} - \frac{1}{V_{ex}}\right)$.

The case of OD group G_5 is the most complicated, since a patron can choose between three route options: ll, lel, and $leel$. By combining the route travel time models (7-9), we find that a patron will choose ll if $\max(X,Y) \leq d_{C_1}^{G_5} \equiv \left(\frac{S_2}{12V_w} + \frac{H_{ex}}{2} + 2\xi_{le} - \xi_{ll}\right)/\left(\frac{1}{V_{lo}} - \frac{1}{V_{ex}}\right)$ and $X + Y \leq d_{C_2}^{G_5} \equiv \left(H_{ex} + \frac{S_1}{2V_{lo}} + 2\xi_{le} + \xi_{ee} - \xi_{ll}\right)/\left(\frac{1}{V_{lo}} - \frac{1}{V_{ex}}\right)$; she will choose $leel$ if $\min(X,Y) \geq d_{C_3}^{G_5} \equiv \left(\frac{H_{ex}}{2} + \frac{5S_1}{12V_{lo}} + \xi_{ee}\right)/\left(\frac{1}{V_{lo}} - \frac{1}{V_{ex}}\right)$ and $X + Y > d_{C_2}^{G_5}$; and will choose lel otherwise.

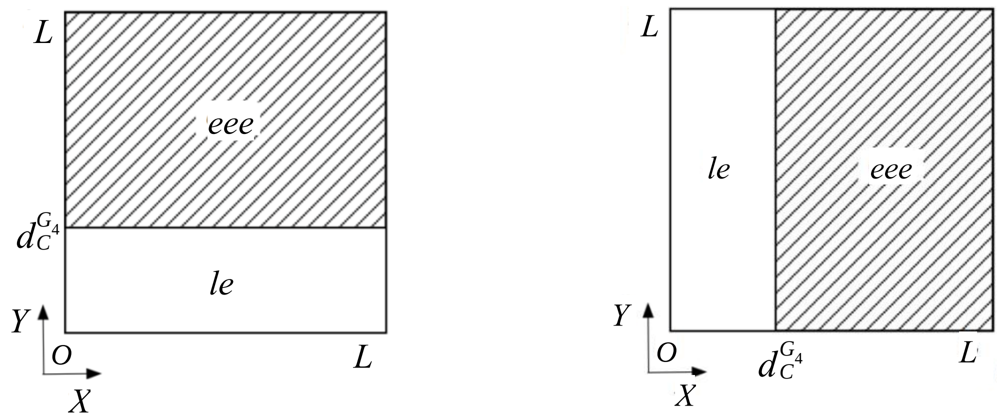
Depending upon the values of $d_{C_1}^{G_5}$, $d_{C_2}^{G_5}$, and $d_{C_3}^{G_5}$, there exist three patterns of partition as shown in Figure C1d.



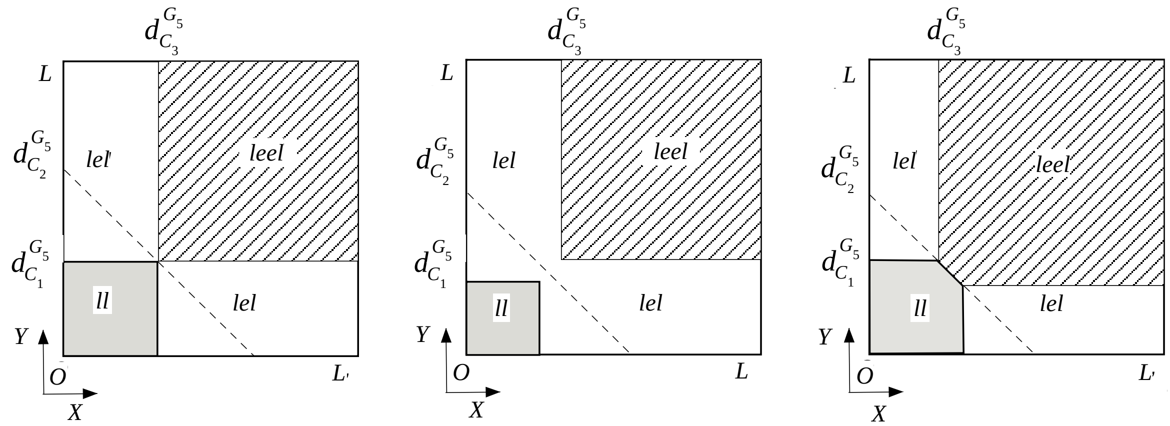
a. OD group G_2, i.e., OD type (1,4)



b. OD group G_3: Left – OD type (2,4), Right – OD type (3,4)



c. OD group G_4: Left – OD type (2,2), Right – OD type (3,3).



d. OD group G_5, i.e., OD type (4,4)

Figure C1. Patrons' route choice regions for OD groups G_2-G_5

The $Q_I^{G_k}$ ($I \in R_k, k = 2,3,4,5$) can thus be calculated by dividing the area of destination region I by the area of the quadrant. The detailed formulas are skipped in this paper for simplicity. Similarly, the expected values of X, Y, $\min(X, Y)$ and $\max(X, Y)$ for a patron in G_k ($k = 2,3,4,5$) who chooses route type I, as well as $E_{I,lo}^{G_k}$ and $E_{I,ex}^{G_k}$ (see Section 2.5), can be calculated as the moments of area for the destination regions in Figures C1a-d. For some complicated cases (e.g. those shown in Figure C1d), the formulas can be derived using an off-the-shelf computer program. In the interest of brevity, these formulas are also skipped in the paper.

Appendix D. Derivation of the critical vehicle occupancies

D.1 For express lines (O_{ex})

We want to show that the patron flow along an express line roughly stays invariant with location. To this end, we consider an *eastbound* express line as shown in Figure D1. We first prove the patron flows are equal between the two sides of a Type-1 stop (i.e. F_1 and F_2 in the figure), and then show that the flows are approximately equal between the two sides of a Type-2 stop (F_3 and F_4 in the figure).

To see why the first half of the above proposition is true, note that for any trip that originates from the Type-1 stop between F_1 and F_2 and takes the eastbound express line first (see the thick dashed line in Figure D1 marked by "trip 1"), there is always a trip ("trip 2" in the figure) that is symmetric to trip 1 about the N-S express line passing through that Type-1 stop. This trip 2 should take the same type of route as trip 1. Moreover, the reverse trip of trip

2 (marked by "trip 3" in the figure) should also take the same type of route, and thus should take the same eastbound express line and end at that Type-1 stop. Therefore, the number of boarding patrons to the eastbound line, who are originated from that Type-1 stop, is equal to the number of alighting patrons who are destined to the same stop. With a similar argument, one can show that the number of patrons who transfer to the eastbound line through that Type-1 stop is also equal to the number who transfer from the same line at the same stop. Hence, the total numbers of boarding and alighting at that Type-1 stop are equal.

Now consider an intermediate express stop between points F_3 and F_4 in Figure D1. An argument similar to the above can be applied to show that the patron flow *approximately* remains the same between F_3 and F_4 if $L \gg S_1$. For a complete proof (which is omitted here for brevity), one needs to verify for every OD type that starts from or ends at that Type-2 stop and takes the eastbound express line, and every OD type that may take a route with a transfer to or from the eastbound express line at that Type-2 stop. For example, one would find that the total number of patrons of OD type (2,1) who board the eastbound line at that Type-2 stop is approximately equal to the number of patrons of OD type (1,2) who alight from the eastbound line at the same stop. To be sure, the network is not perfectly symmetric about the vertical axis passing through the Type-2 stop, but the difference between the left and right sides of that stop is small when $L \gg S_1$ (which is guaranteed by constraint (1h) in Section 2.2). This means the above approximation should be good.

Therefore, an express vehicle's patron occupancy remains approximately constant along its journey; and the critical occupancy, O_{ex}, is equal to the average occupancy along the line, which is given by (21).

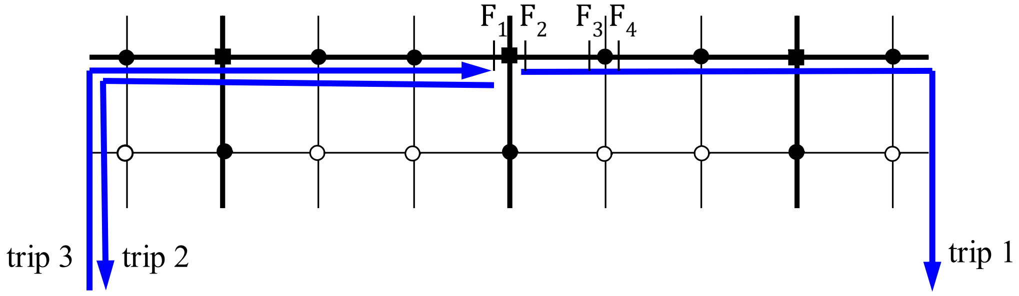


Figure D1. Patron flows on an eastbound express line

D.2 For local lines (O_{lo})

Unlike the express lines, the patron flow passing through a local line varies significantly with the location. We first identify the locations where the maximum flow occurs, and then formulate that maximum flow.

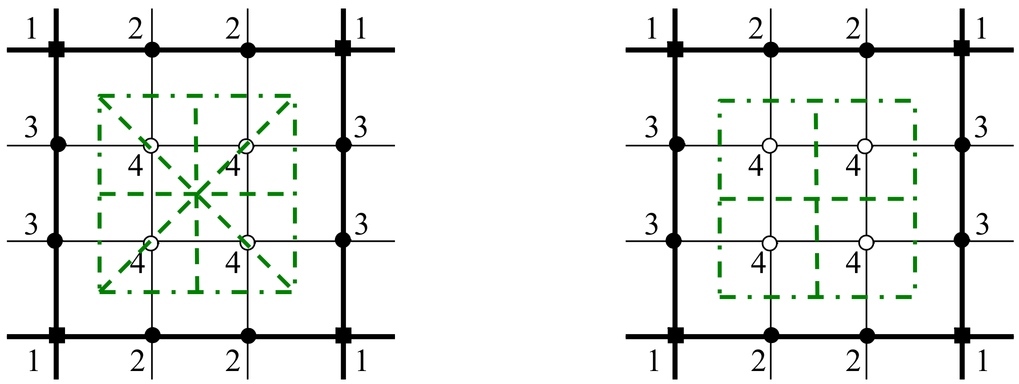
First note that all the local travel segments that may occur in the intersecting network can be divided into two classes: i) direct travel segments, each of which covers the entire distance in one travel direction (E-W or N-S); and ii) feeder segments. The local travel segments in route types ll, le, and lel are direct segments; and those in route types lee and $leel$ are feeder ones (see again Figures 4b-e). For the direct local segments, the method presented in Appendix D.1 can be applied to show that the resulting patron flow (which is a portion of the total flow on a local line) is approximately constant along the local lines given $L \gg S_1$.

The patron flow contributed by the feeder segments, on the other hand, is unevenly distributed along a local line. Since a feeder segment can be considered as a many-to-one system with evenly distributed trip origins, the maximum patron flow is always attained at its end, i.e., at the local line sides of Type-2 and 3 stops. Further, this maximum flow for a specific feeder segment is twice of the average patron flow along this segment. Given the above, we now examine the maximum patron flows of these feeder segments by OD types.

(i) For OD group G_1 (OD type (1,4)) and route type lee, a patron always takes the local service to access the nearest express line, regardless of the direction. Figure D2a shows a type-4 zone of trip origins (see also Figure 5) enclosed by dash-dot lines for a case of $m = \frac{S_1}{S_2} = 3$. The zone is divided into eight parts by the dashed lines. Trips originated in each part will feed into the nearest Type-2 or 3 stop. Thus, the maximum local-line patron flow for this OD type is attained at a Type-2 or 3 stop that is the nearest to the mid-point between two neighboring Type-1 stops, because that stop has the largest catchment area as a feeder destination. (Specifically, in Figure D2a with $m = 3$, all the Type-2 and 3 stops have the same maximum local patron flow.) For this type of feeder segments, we can calculate by geometry the ratio between the maximum patron flow and the average patron flow across all the feeder segments, denoted by ρ_o. Note that this ratio depends only on m. The formula of ρ_o is furnished in (23) while the detailed derivation is omitted for brevity.

(ii) For OD type (2,4) and route type lee, a patron will take an E-W local line as feeder to access the nearest Type-3 stop. Figure D2b illustrates the catchment areas of the Type-3 stops for this type of OD. Similarly, a patron of OD type (3,4) and route type lee will always access the nearest Type-2 stop via a N-S local line. In both cases, the sizes of the catchment areas are equal and the maximum local-line patron flow is attained at the sides of all the Type-2 or 3 stops. Since the demands of OD types (2,4) and (3,4) are also equal, the maximum patron flow contributed by the feeder segments of OD group G_3 should be twice of the average flow.

(iii) For OD group G_5 (OD type (4,4)) and route type $leel$, a patron can freely choose between the horizontal and vertical local lines as the feeder on either end of her trip, but the directions of the two feeder segments have to be perpendicular. Hence, the ratio between the maximum patron flow and the average flow is $\frac{\rho_o+2}{2}$ for this case.



a. For OD type (1,4) and route type lee b. For OD type (2,4) and route type lee

Figure D2. Catchment areas of feeder destinations (Type-2 and 3 stops)

Combining all the above cases for both direct and feeder segments of local travel, we conclude that the maximum patron flow of the local network would occur at the local sides of a Type-2 or 3 stop that is the closest to the midpoint between two neighboring Type-1 stops. The resulting critical vehicle occupancy, O_{lo}, is formulated as follows:

$$O_{lo} = \frac{\lambda}{K_{lo}} \sum_{k=2}^{5} \sum_{I \in R_k} P_{G_k} Q_I^{G_k} E_{I,lo}^{G_k} \rho_{I,lo}^{G_k} \tag{D1}$$

where $\rho_{I,lo}^{G_k}$ ($I \in R_k, k = 2,3,4,5$) denotes the ratio between the maximum patron flow and the average flow associated with the local segments of route type I for OD group G_k. Specifically, $\rho_{le,lo}^{G_2} = \rho_{le,lo}^{G_3} = \rho_{le,lo}^{G_4} = \rho_{ll,lo}^{G_5} = \rho_{lel,lo}^{G_5} = 1$ because they are direct local segments, $\rho_{lee,lo}^{G_2} = \rho_o$ (above case (i)), $\rho_{lee,lo}^{G_3} = 2$ (above case (ii)), $\rho_{leel,lo}^{G_5} = \frac{\rho_o+2}{2}$ (above case (iii)), and $\rho_{eee,lo}^{G_4} = 0$. This is exactly (22) in Section 2.5.

References

APTA, 2007. A Profile of Public Transportation Passenger Demographics and Travel Characteristics. Washington, DC: American Public Transportation Association.

Badia, H., Estrada, M., Robusté, F., 2014. Competitive transit network design in cities with radial street patterns. *Transportation Research Part B* 59, 161-181.

Byrne, B.F., 1975. Public transportation line positions and headways for minimum user and system cost in a radial case. *Transportation Research* 9 (2), 97-102.

Ceder, A., Wilson, N., 1986. Bus network design. *Transportation Research Part B* 20, 331-344.

Chang, S.K., Schonfeld, P., 1993. Optimal dimensions of bus service zones. *Journal of Transportation Engineering* 119(4), 567-585.

Chen, H., Gu, W., Cassidy, M.J., Daganzo, C.F., 2015. Optimal transit service stop ring-radial and grid street networks: a continuum approximation design method and comparisons. *Transportation Research Part B* 81, 755-774.

Chen, P.W., Nie, Y.M., 2017a. Analysis of an idealized system of demand adaptive paired-line hybrid transit. *Transportation Research Part B* 102, 38-54.

Chen, P.W., Nie, Y.M., 2017b. Connecting e-hailing to mass transit platform: Analysis of relative spatial position. *Transportation Research Part C* 77, 444-461.

Chen, P.W., Nie, Y.M., 2018. Optimal design of demand adaptive paired-line hybrid transit: Case of radial route structure. *Transportation Research Part E* 110, 71-89.

Chien, S., Schonfeld, P., 1997. Optimization of grid transit system in heterogeneous urban environment. *Journal of Transportation Engineering* 123(1), 28-35.

Chien, S., Schonfeld, P., 1998. Joint optimization of a rail transit line and its feeder bus system. *Journal of Advanced Transportation* 32(3), 253-284.

Daganzo, C.F., 2010a. Structure of competitive transit networks. *Transportation Research Part B* 44, pp. 434-446.

Daganzo, C.F., 2010b. Public Transportation Systems: Basic Principles of System Design, Operations Planning and Real-Time Control. University of California, Berkeley.

Estrada, M., Roca-Riu, M., Badia, H., Robusté, F., Daganzo, C., 2011. Design and implementation of efficient transit networks: procedure, case study and validity test. *Transportation Research Part A* 45, 935-950.

Fan, W., Machemehl, R., 2004. Optimal Transit Route Network Design Problem: Algorithms, Implementations, and Numerical Results. *Technical Report* SWUTC/04/167244-1. University of Austin. Texas, USA.

Fan, W., Mei, Y., 2018. Joint design of trunk-feeder transit with many-to-many demand. *The 97th Annual Meeting of Transportation Research Board*, Washington, DC.

Fan, W., Mei, Y., Gu, W., 2017. Design of bimodal local and express transit service on a linear corridor with many-to-many demand. *The 22nd International Conference of Hong Kong Society for Transportation Studies*.

Freyss, M., Giesen, R., Munoz, J.C., 2013. Continuous approximation for skip-stop operation in rail transit. *Transportation Research Part C* 36, 419-433.

Gu, W., Amini, Z., Cassidy, M.J., 2016. Exploring alternative service schemes for busy transit corridors. *Transportation Research Part B* 93, 126-145.

Guan, J., Yang, H., Wirasinghe, S.C., 2003. Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B* 40, 885-902.

Guo, Z., 2008. Transfers and Path Choice in Urban Public Transport Systems. *Master Dissertation*, Massachusetts Institute of Technology.

Guo, Z., Wilson, N.H.M., 2004. Assessment of the transfer penalty for transit trips: Geographic information system–based disaggregate modeling approach. *Transportation Research Record* 1872, 10-18.

Holryod, E., 1967. Optimum bus service: A theoretical model for a large uniform urban area. In: Edie, L. (Ed.), *Proceedings of the Third International Symposium on the Theory of Traffic Flow*. Elsevier, 308-328.

Hurdle, V.F., Wirasinghe, S.C., 1980. Location of rail stations for many to one travel demand and several feeder modes. *Journal of Advanced Transportation* 14(1), 29-45.

Ibarra-Rojas, O.J., Delgado F., Giesen, R., Muñoz, J.C., 2015. Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B* 77, 38-75.

Kepaptsoglou, K., Karlaftis, M.G. 2009. Transit route network design problem: Review. *Journal Transportation Engineering* 135 (8), 491-505.

Li, X., Quadrifoglio, L., 2011. 2-Vehicl zone optimal design for feeder transit services. *Public Transportation* 3, 89-104.

Liu, R., Pendyala, R.M., Polzin, S., 1997. Assessment of intermodal transfer penalties using stated preference data. *Transportation Research Record* 1607, 74-80.

Newell, G.F., 1973. Scheduling, location, transportation, and continuum mechanics: some simple approximations to optimization problems. *SIAM Journal on Applied Mathematics* 25(3), 346-360.

Newell, G.F., 1979. Some issues relating to the optimal design of bus routes. *Transportation Science* 13(1), 20-35.

Newell, G.F., 1994. Lecture Note on Public Transportation. University of California, Berkeley.

Ouyang, Y., Nourbakhshet, S.M., Cassidy, M.J., 2014. Continuum approximation approach to bus network design under spatially heterogeneous demand. *Transportation Research Part B* 68, 333-344.

Saidi, S., Wirasinghe, S.C., Kattan, L., 2016. Long-term planning for ring-radial urban rail transit networks. *Transportation Research Part B* 86, 128-146.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. New Jersey: Prentice-Hall, Inc.

Sivakumaran, K., Li, Y., Cassidy, M.J., Madanat, S., 2012. Cost-saving properties of schedule coordination in a simple trunk-and-feeder transit system. *Transportation Research Part A* 46, 131-139.

Sivakumaran, K., Li, Y., Cassidy, M., Madanat, S., 2014. Access and the choice of transit technology. *Transportation Research Part A* 59, 204-221.

Vuchic, V.R., 2007. Urban Transit System and Technology. Hoboken, NJ: John Wiley & Sons.

Vuchic, V.R., Newell, G.F., 1968. Rapid Transit interstation spacings for minimum travel time. *Transportation Science* 2(4), 303-339.

Wan, Q., Lo, H., 2003. A mixed integer formulation for multiple-route transit network design. *Journal of Mathematical Modelling and Algorithms* 2, 299-308.

Wan, Q., Lo, H., 2009. Congested multimodal transit network design. *Public Transportation* 1, 233-251.

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers Part II* 1, 325-378.

Wirasinghe, S.C., 1980. Nearly optimal parameters for a rail/feeder-bus system on a rectangular grid. *Transportation Research Part A* 14A, 33-40.

Wirasinghe, S.C., 2003. Initial Planning for Urban Transit Systems. In *Advanced Modeling for Transit Operations and Service Planning* edited by Lam W.H.K. and Bell M.G.H., Oxford, UK: Pergamon.

Wirasinghe, S.C., Ghoneim, N.S., 1981. Spacing of bus stops for many to many travel demand. *Transportation Science* 15 (3), 210-221.

Wirasinghe, S.C., Hurdle, V., Newell, G., 1977. Optimal parameters for a coordinated rail and bus transit system. *Transportation Science* 11, 359-374.