Network Design for Personal Rapid Transit under Transit-Oriented Development

Hong Zheng\textsuperscript{a}, Srinivas Peeta\textsuperscript{b}\textsuperscript{*}

\textsuperscript{a} NEXTRANS Center, Purdue University, 3000 Kent Ave., West Lafayette, IN 47906, USA
\textsuperscript{b} School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907, USA

Submitted for publication in Transportation Research Part C: Emerging Technologies: "Optimization (OPT-i)" – Professor Matthew G. Karlaftis Memorial Issue

\textsuperscript{*}Corresponding author. Tel.: +1 765 494 2209; fax: +1 765 496 7996.
E-mail addresses: zheng225@purdue.edu (H. Zheng), peeta@purdue.edu (S. Peeta)
ABSTRACT

This study addresses guideway network design for personal rapid transit (PRT) favoring transit-oriented development. The guideway network design problem seeks to minimize both the guideway construction cost and users’ travel time. In particular, a set of optional points, known as Steiner points, are introduced in the graph to reduce the guideway length. The model is formulated as a combined Steiner and assignment problem, and a Lagrangian relaxation based solution algorithm is developed to solve the optimal solution. Numerical studies are carried on a real-sized network, and illustrate that the proposed model and solution algorithm can solve the PRT guideway network design problem effectively.

Keywords: network design, Steiner problem, Lagrangian relaxation, personal rapid transit, transit-oriented development
1. Introduction

Personal rapid transit (PRT) is an automated transit system in which vehicles are used to transport a batch of passengers on its own right-of-way on demand to their destinations without stops and transfers (Anderson, 1998). A PRT system provides a service similar to taxi because passengers are served on demand and there are no pre-determined schedules for PRT. At the PRT station, a group of passengers first select the intended destination station. A PRT vehicle is then dispatched to the station to carry the passengers to the desired destination. Stations are offline such that vehicles can accelerate/decelerate on auxiliary lanes without interfering with the vehicles passing by the main through-lanes. Hence, a PRT vehicle can operate in a non-stop manner by bypassing all intermediate stations. Currently, PRT vehicles supported by modern technologies are usually designed to run on electricity, and are operated by computer control requiring no driver. In general, the size of a PRT vehicle can accommodate three to six passengers.

PRT vehicles run exclusively on tracks called guideways. The guideways are designed as facilities to eliminate at-grade crossings or interference with other transportation modes. In the U.S., PRT has been implemented as a mode of automated people movers at airports and institutions such as schools (for example, the PRT system in West Virginia University campus in Morgantown, WV (Sproule and Neumann, 1991)). Worldwide, PRT systems have been considered in several real-world applications recently, including in Korea (Suh, 2001), Sweden (Tegner et al., 2007) and United Arab Emirates (Mueller and Sgouridis, 2011).

In recent planning practice for urban development in the future, there has been an increasing emphasis in the global community on sustainable transportation systems. The excessive use of personal cars has been identified with issues related to congestion, energy consumption, air pollution, noise, safety and excessive land use. Transit-oriented development (TOD) has emerged as a promising alternative for sustainable communities to overcome the aforementioned issues by creating compact environments using convenient and efficient public transportation systems. TOD is deployed to reduce people’s dependence on personal cars for mobility and to help make livable and vibrant communities. The most vital element in a TOD design is the planning and design of the public transportation network which serves as the backbone of urban infrastructure systems (Lin and Shin, 2008, Li et al., 2010). A recent trend in TOD deployment is to introduce efficient transit systems such as bus rapid transit (BRT), light rail transit (LRT) and personal rapid transit (PRT) (Parent, 2006). Among these modes, PRT has received significant attention as it connects personal, private, and public transportation, and because of its flexible operational characteristics and competitive financial aspects (Muir et al., 2009). For instance, Tegner et al. (2005) estimated that the construction cost of a
The PRT system is about a third that of light rail, as PRT entails much lighter vehicle size and lower design standards for guideways compared to LRT.

Mass transportation systems such as trains, metros and rapid buses constitute the major modes of TOD development (see Figure 1). These modes are efficient for transporting passengers, measured in terms of per unit of space or energy, provided that the demand is adequate. However, if the demand decreases, the ridership drops while the operational cost remains the same, degrading the system efficiency. This is the key reason why most mass transit systems reduce frequencies during off-peak hours (Clerget et al., 2001). Therefore, each mass transit system has a certain operating range in terms of passengers per hour to maintain an efficient operation of the system. To facilitate TOD development, an alternative to the personal car needs to provide a public transit mode which offers the same door-to-door flexibility at an acceptable cost. This could be achieved through a mixed design combining high passenger-flows mass transit and flexible public transportation carrying low passenger-flows for those times or places. PRT is one such flexible system serving as a supplementary mode for TOD development, where a PRT system functions as a local area network, connecting the traditional transit systems and other transit modes within its network.

(Place Fig. 1 here)

PRT could be a sustainable solution to urban problems as well. Congestion in major cities results in not only severe travel time delay, but also excess energy use and emissions. PRT is a potential solution to reduce congestion on urban highways. Further, because a PRT system is electrically powered, there are no emissions, and thus overall energy and emissions could be significantly reduced. Compared with automobiles, the energy savings for a PRT vehicle could be about 75%, and CO2 emissions reduction could be more than 60% (Lowson, 2003a). Land use area is also reduced because of the small footprint of the system compared to the traditional road infrastructure.

While PRT has been recognized as an important component of alternate solutions to passenger cars in sustainable transportation systems in the future, it has not yet achieved widespread commercial deployment in the U.S. Two major factors that restrict the PRT deployment in practice are cost and line capacity. Studies show that the construction cost of guideways is estimated between $5–$15 million per lane per mile; of the stations about $0.5–$3 million per station, and of the vehicles about $0.2–$0.7 million per vehicle (in 1996 dollars) (Yoder et al., 2000). For example, the PRT system at West Virginia University in Morgantown consists of 8.7 miles of guideways and 5 stations, and cost over $126 million in 1979 – about $319 million in 2004 dollars (Sproule and Neumann, 1991). The majority of the cost is for the guideway construction. Hence, the reduction of the guideway length is critical to reduce the PRT construction cost.
The PRT may also have a limited line capacity as compared to other public transportation systems. The line capacity is governed by the allowable vehicle headways, which are further dictated by the safety requirement for a brick-wall stop. Since PRT is designed to provide flexible door-to-door accessibility for a small group of passengers, the size of a PRT vehicle is small, and accommodates up to 6 passengers. The current design speed of PRT is also relatively low, based not only on passenger comfort but also on the low design standard for guideways to reduce the construction costs. For instance, the design speed of the Morgantown PRT system is 30 mile per hour (Juster and Schonfeld, 2013). Further, to meet safety standards, the headway ranges from 8 to 15 seconds in current practice (Juster and Schonfeld, 2013). These factors lead to a limited line capacity of around 2,000 – 2,500 passengers per hour, which is less than that of the conventional public transit modes with larger fleet size and higher operating speed.

The limitations associated with both cost and line capacity could be improved through appropriate guideway network (GN) design, because a well-designed GN can improve connectivity and accessibility while providing more options for route choice. Unlike other public transportation modes that generally operate through a single line, a PRT system is usually designed as an interconnected system (or grid of guideways) with junctions. Since vehicles do not have to follow a pre-defined route, the system allows a PRT vehicle to flexibly select a route from a variety of routes in the network. Thus, the overall throughput of the entire network could be improved (Carnegie and Hoffman, 2007).

Past PRT-related research has focused on network design (Won et al., 2006, Ma and Schneider, 1991, Won and Karray, 2008, Kornhauser, 2009), capacity analysis (Lees-Miller et al., 2010, Schweizer and Mantecchini, 2007, Lowson, 2003b, Mueller and Sgouridis, 2011, Johnson, 2005), and empty vehicle management to reduce passenger wait time (Andreasson, 1994, Schweizer et al., 2012, Andreasson, 2003, Lees-Miller et al., 2010). In this paper, we analyze the PRT GN design to support TOD deployment. That is, the model is multimodal and incorporates vehicle and transit networks and their interconnections, and integrates with other modes of public transit, such as commute bus, light rail, heavy rail, metro system, etc. The GN design involves two objectives that may conflict with each other. The first objective is from the supplier perspective. As the construction cost is proportional to the GN length, there is a need to minimize the total GN length. The second objective arises from the user standpoint. As an efficient GN system needs to reduce the travel time, there is a need to minimize the total in-vehicle travel time across all users. Since these two objectives can conflict with each other, it leads to a bi-criteria network design problem, and the optimal solution is pareto-optimal. In this study, we assign different weights to the two objectives to balance the two criteria,

---

1 Some studies argue that the headway could ideally be reduced to as low as 3 seconds. However, this has not been used in real-world PRT applications.
and solve the optimal solution for the weighted objective function rather than solving for the pareto-optimal solution.

The remainder of the paper is organized as follows. Section 2 briefly discusses the overall methodology, and specifically focuses on the theoretical approach related to reducing the GN length. Section 3 proposes a multi-commodity flow formulation for the GN network design, and a solution algorithm based on Lagrangian relaxation is discussed in Section 4. Section 5 presents a case study to demonstrate the effectiveness of the proposed methodology. Section 6 provides some concluding comments.

2. Methodology

Given a set of PRT stations (also called terminals), denoted by $V$, and the origin-destination (O-D) demand between the stations, the goal of the GN design is to establish a graph $G = (N, A)$, where $N$ denotes a set of nodes and $A$ denotes a set of arcs (also called links), connecting the PRT stations $V$ to address to two objectives: (1) minimize the total link length; and (2) minimize the total system travel time experienced by travelers. The first objective is intended to minimize the construction and maintenance cost of the GN, and the second objective is to minimize travel time and improve the level of service for users.

2.1. Reducing the GN length

In the endeavor to reduce the network length, we show that a set of intermediate points that may not be strictly needed to design the connecting network (GN) could be very useful in reducing the length. In graph theory, such intermediate points optional in network connectivity are known as the Steiner points, and the problem is known as the Steiner problem on the graph. If the connecting network is restricted to a tree, the GN design subject to the minimum arc length is known as the min-cost Steiner tree problem (STP), which is NP-hard (Dreyfus and Wagner, 1971, Winter, 1987, Hakimi, 1971). Figure 2 shows an example to illustrate how Steiner points can reduce the length in the GN design. More specifically, Figure 2(a) indicates that the minimum length required to connect the three stations is 6, where the distance between each pair of stations is 2. Figure 2(b) illustrates that by introducing an intermediate (Steiner) point in the middle, which is optional in connecting the three stations, the length could be halved.

(Place Figure 2 here)

The PRT GN design in this paper leverages the concept of Steiner points. Note that the PRT GN design in our problem is not restricted to a tree. Hence, the model formulated in Section 3 will exhibit the Steiner point feature (i.e., optional connecting points) but the problem is not a STP. Further, since PRT guideways are subject to the line capacity representing the maximum vehicles per hour that can be afforded by the PRT service,
arcs in the network must have capacity – another component differentiating our model from the STP.

2.2. Reducing the system travel time

For the PRT GN design, we introduce another objective, from the users’ perspective, which is to reduce the travel time of users. As a PRT vehicle travels at a constant speed in general, the PRT vehicle routing policy is the shortest distance (or travel time). Each link \((i,j) \in A\) is associated with a constant traversal time \(t_{ij}\), and capacity \(u_{ij}\), representing the maximum hourly vehicular flow that can pass that link. We then assign the O-D demand in the GN such that the system travel time is minimized while satisfying the capacity constraint. This is a multi-commodity flow assignment problem.

In summary, the PRT GN design problem could be understood as a combined Steiner problem on the graph and min-cost multi-commodity flow assignment problem, as formulated in the next section.

3. Model formulation

3.1. Notation

A PRT GN \(G = (N, A)\) consists of a set of nodes \(N\) and a set of directed arcs \(A\). A set of terminals (PRT stations), denoted by \(V \subseteq N\), must be connected by \(A\). A set of nodes \(S = N - V\) is the set of Steiner points. We assume that the geographical locations of the nodes in \(S\) are given. Let \(c_{ij}\) denote the length, \(t_{ij}\) the travel time, and \(u_{ij}\) the capacity of arc \((i,j) \in A\), respectively. Let the flow assigned on arc \((i,j)\) be denoted by \(x_{ij}\). Let the link construction decision be denoted by a binary variable \(y_{ij} \in \{0,1\}\); \(y_{ij} = 1\) if link \((i,j)\) is constructed, \(y_{ij} = 0\) otherwise.

Let \(P\) denote a set of origins and \(Q\) a set of destinations. A set of origin-destination (O-D) pairs is denoted by a vector \(K := \{(p,q) \mid p \in P, q \in Q\}\). Let \(k \in K\) denote a commodity associated with the O-D pair \((p,q)\). Let \(p(k)\) represent the origin of commodity \(k\) and \(q(k)\) represent the destination of commodity \(k\). Let \(b^k\) be the corresponding O-D demand and \(x_{ij}^k\) the amount of commodity \(k\) assigned on arc \((i,j)\).

3.2. Multi-commodity flow formulation

The PRT GN problem is formulated as a multi-commodity flow problem as follows.

\[
\min_{x,y} \alpha \sum_{(i,j) \in A} c_{ij} \cdot y_{ij} + \beta \sum_{k \in K} \sum_{(i,j) \in A} x_{ij}^k \cdot t_{ij}
\]
The first term of the objective function (Eq.(1a)) is to minimize the total guideway length; that is, to minimize the construction (and maintenance) cost of guideways. The second term is to minimize the total travel time over all users. If $\beta = 0$, then the problem becomes a min-cost Steiner tree problem; if $\alpha = 0$, the problem then becomes a multi-commodity min-cost flow problem. The overall objective function is balanced by assigning different weights, $\alpha$ and $\beta$, to the first and second terms, respectively. Different values of $\alpha$ and $\beta$ will result in different network topologies in the solution. Eq.(1b) is flow mass balance constraint for each commodity $k$. Eq.(1c) indicates that a commodity on arc $(i,j)$ can be positive only if $(i,j)$ is constructed (i.e., $y_{ij} = 1$), and it is zero if $(i,j)$ is not constructed (i.e., $y_{ij} = 0$). Note that the amount of commodity $k$ on an arc $(i,j)$ is at most $b_k^i$. Constraints (1b)-(1c) indicate that a feasible solution must have a directed path of arcs (i.e., $y_{ij} = 1$) for each commodity $k \in K$. Thus, we model the network connectivity via an embedded multi-commodity network flow problem. Eq.(1d) is the capacity constraint as the PRT guideway is in general subject to a maximum service rate. Hence, the commodity flow on arc $(i,j)$ should be less than the capacity $u_{ij}$. If the problem is uncapacitated, we assign $u_{ij}$ a sufficiently large number. To maintain connectivity between the terminal set $V$, Eq.(1e) indicates that the number of constructed arcs is bounded between $[M_1,M_2]$. To connect $|V|$ terminals it requires at least $|V| - 1$ arcs (i.e., terminal $V$ is connected by a tree in a complete graph); so, the lower bound of $M_1$ is $|V| - 1$. At most all arcs are constructed; so, the upper bound of $M_2$ is $|A|$. Therefore, initially one can set $M_1 := |V| - 1$, and $M_2 := |A|$. Eq.(1e) initially is redundant, however, it is shown in the solution algorithm discussed in Section 4 that this range could be tightened iteratively in the algorithmic procedure and thus can provide a better bound (Beasley, 1984). Eq.(1f) specifies flow non-negativity, and Eq.(1g) specifies a binary variable for $y_{ij}$.  

\[
\sum_{(i,j) \in A} x_{ij}^k - \sum_{(j,i) \in A} x_{ji}^k = \begin{cases} b_k^i & i = p(k) \\ -b_k^i & i = q(k) \\ 0 & \text{otherwise} \end{cases} \quad \forall k \in K, \forall i \in N \quad (1b) \\

x_{ij}^k \leq y_{ij} \cdot b_k^i \quad \forall k \in K, \forall (i,j) \in A \quad (1c) \\

\sum_{k \in K} x_{ij}^k \leq u_{ij} \quad \forall (i,j) \in A \quad (1d) \\

M_1 \leq \sum_{(i,j) \in A} y_{ij} \leq M_2 \quad (1e) \\

x_{ij}^k \geq 0 \quad \forall k \in K, \forall (i,j) \in A \quad (1f) \\

y_{ij} \in \{0,1\} \quad \forall (i,j) \in A \quad (1g)
To reduce the size of the feasible set, we reformulate the model with the narrowed solution space, called a strong formulation. We denote by \( f_{ij}^k \) the proportion of commodity \( k \) assigned on arc \((i,j)\), i.e., \( f_{ij}^k = \frac{x_{ij}^k}{b_k} \). The new variable \( f_{ij}^k \) has \( f_{ij}^k \in [0,1] \) and its feasible region is much smaller than the original variable \( x_{ij}^k \in [0,b_k] \). Because commercial Integer Programming (IP) solvers usually use the branch-and-bound algorithm solve an IP, the tighter feasible set in a strong formulation can lead to fewer branch-and-bound variables. The problem can then be reformulated as follows.

\[
\begin{align*}
\min \quad & \alpha \sum_{(i,j) \in A} c_{ij} \cdot y_{ij} + \beta \sum_{k \in K} \left( \sum_{(i,j) \in A} f_{ij}^k \cdot t_{ij} \cdot d^k \right) \\
\text{s.t.} \quad & \sum_{(i,j) \in A} f_{ij}^k - \sum_{(j,i) \in A} f_{ij}^k = \begin{cases} 1 & i = p(k) \\ -1 & i = q(k) \\ 0 & \text{otherwise} \end{cases} \quad \forall k \in K, \forall (i,j) \in A \quad (2b) \\
& f_{ij}^k \leq y_{ij} \quad \forall k \in K, \forall (i,j) \in A \quad (2c) \\
& \sum_{k \in K} f_{ij}^k \cdot d^k \leq u_{ij} \quad \forall (i,j) \in A \quad (2d) \\
& 0 \leq f_{ij}^k \leq 1 \quad \forall k \in K, \forall (i,j) \in A \quad (2e) \\
\end{align*}
\]

(1e, 1g)

4. Lagrangian-relaxation based solution algorithm

4.1. Lagrangian relaxation

In the formulated problem (2), Eqs.(1e) and (1g) only involve decision variable \( y_{ij} \), which is the network connectivity component. Eqs.(2b), (2d) and (2e) only involve decision variable \( f_{ij}^k \), which is the flow assignment component. Note that Eq.(2c) can be rewritten to an aggregate constraint in Eq.(3). In the solution algorithm, as Eq.(2c) defines a tighter feasible region than the aggregated constraint (3), the former provides a better bound, which is a desirable feature in the algorithmic development. Therefore, we use the disaggregate constraint (2c) instead of (3) in the PRT model formulation.

\[
\sum_{k \in K} f_{ij}^k \leq |K| \cdot y_{ij} \quad \forall (i,j) \in A \quad (3)
\]

Eq.(2c) is the only coupling constraint which couples decisions variable \( f_{ij}^k \) and binary variable \( y_{ij} \); thus it is the hard constraint coupling the flow assignment and network connectivity components, which makes problem (2) hard to solve. The main idea of Lagrangian relaxation is to relax the hard constraints and place them in the objective
function. As such, we dualize the hard constraint (2c) by associating a nonnegative Lagrangian multiplier (or dual variable vector) $\lambda = (\lambda_{ij}^k)_{k \in K, (i,j) \in A} \geq 0$. The Lagrangian dual problem ($LD$) is then formulated in (4). The constraint set of $LD$ is same as problem (2), except constraint (2c) which has been dualized in the objective function.

$$LD: \max_{\lambda} \min_{t, y} \sum_{(i,j) \in A} \left( \alpha \cdot c_{ij} - \sum_{k \in K} \lambda_{ij}^k \right) \cdot y_{ij} + \sum_{k \in K} \sum_{(i,j) \in A} \left( \beta \cdot t_{ij} \cdot d^k + \lambda_{ij}^k \right) \cdot f_{ij}^k$$ (4)

s.t. 

$$(1e, 1g, 2b, 2d, 2e)$$

We seek to solve $LD$ instead of solving the primal problem directly. The goal is to solve for the Lagrange multiplier $\lambda \geq 0$ that maximizes $LD$. With a given feasible $\lambda$, $LD$ provides a lower bound due to the Lagrangian duality theorem. Before we discuss how to solve $\lambda$ that maximizes $LD$, we first consider the following relaxation problem. Suppose $\lambda$ is given and fixed, then the problem of $LD$ is relaxed to problem (5). The relaxed problem is also called the Lagrangian relaxation problem ($LR$) with parameter $\lambda \geq 0$. The general idea is to solve a set of $LR$ problems, each one is with a given and fixed $\lambda$ which provides a lower bound, and then find $\lambda$ that has the tightest bound. With a fixed $\lambda$, $LR$ can be easily solved, because the hard constraint (2c) is relaxed and placed in the objective function. Next, we discuss how to solve $LR$ in a straightforward manner.

$$LR: \min_{t, y} \sum_{(i,j) \in A} \left( \alpha \cdot c_{ij} - \sum_{k \in K} \lambda_{ij}^k \right) \cdot y_{ij} + \sum_{k \in K} \sum_{(i,j) \in A} \left( \beta \cdot t_{ij} \cdot d^k + \lambda_{ij}^k \right) \cdot f_{ij}^k$$ (5)

s.t. 

$$(1e, 1g, 2b, 2d, 2e)$$

By inspection, problem $LR$ is separable, i.e., it can be separated into two subproblems, $LR(1)$ and $LR(2)$, which are independent of each other. $LR(1)$ is the network connectivity subproblem, and $LR(2)$ is the flow assignment subproblem. Next, we show that both problems could be solved trivially.

$$LR(1): \min_{y} \sum_{(i,j) \in A} \left( \alpha \cdot c_{ij} - \sum_{k \in K} \lambda_{ij}^k \right) \cdot y_{ij}$$ (6a)

$$M_1 \leq \sum_{(i,j) \in A} y_{ij} \leq M_2$$ (6b)

s.t. 

$$y_{ij} \in \{0,1\} \quad \forall (i,j) \in A$$ (6c)

$$LR(2): \min_{t} \sum_{k \in K} \sum_{(i,j) \in A} \left( \beta \cdot t_{ij} \cdot d^k + \lambda_{ij}^k \right) \cdot f_{ij}^k$$ (7a)
\[ \sum_{(i,j) \in A} f_{ij}^k - \sum_{(j,i) \in A} f_{ji}^k = \begin{cases} 1 & i = p(k) \\ -1 & i = q(k) \\ 0 & \text{otherwise} \end{cases} \quad \forall k \in K, \forall i \in N \quad (7b) \]

\[ \sum_{k \in K} f_{ij}^k \cdot d^k \leq u_{ij} \quad \forall (i,j) \in A \quad (7c) \]

\[ 0 \leq f_{ij}^k \leq 1 \quad \forall k \in K, \forall (i,j) \in A \quad (7d) \]

\( LR(1) \) is an integer program, and can be easily solved by inspection. Denote a composite cost \( c'_{ij} = \alpha \cdot c_{ij} - \sum_{k \in K} \lambda^k_{ij} \). We rank arcs in \( A \) in the increasing order of \( c'_{ij} \). Let \( r_{ij} \) represent the order of an arc \( (i,j) \). The optimal solution of \( LR(1) \) is:

\[ y_{ij} = \begin{cases} 1 & \text{if } r_{ij} \leq M_1 \\ 1 & \text{if } c'_{ij} \leq 0 \text{ and } M_1 < r_{ij} \leq M_2 \\ 0 & \text{otherwise} \end{cases} \quad (8) \]

Eq.(8) indicates that \( y_{ij} = 1 \) for the first \( M_1 \) arcs with the smallest \( c'_{ij} \), and \( y_{ij} = 1 \) for the remaining arcs which have \( c'_{ij} \leq 0 \). Note that Eq.(8) solves an integer solution directly, and therefore our solution algorithm avoids the need for branch-and-bound.

\( LR(2) \) can also be easily solved as it is simply a multi-commodity network flow problem, which is a linear program (LP) and can be solved by any commercial LP solver.

4.2. Solving the Lagrangian dual problem

The sub-gradient approach (Held et al., 1974) is a common method to solve the Lagrangian dual problem \( LD \). The idea is to generate a sequence \( \lambda^0, \lambda^1, ..., \lambda^n \) of Lagrange multiplier vectors heuristically following a descent direction. At each iteration, we solve the Lagrangian relaxation problem \( LR \) by solving \( LR(1) \) and \( LR(2) \) separately. Eq.(5) then provides an optimal solution of \( LR \) for a given \( \lambda \), which is a lower bound to the primal problem denoted by \( Z_{LB} \). Based on the solution of \( f \) obtained by solving \( LR(2) \), we can also obtain a feasible solution of \( y \) to the primal problem trivially, by setting \( y_{ij} = 1 \) if \( (i,j) \in S \), where \( S \) denotes the arc set \( S = \{(i,j) | f_{ij}^k > 0, \forall (i,j) \in A, k \in K\} \). It implies that if an arc \( (i,j) \) carries positive flows for a commodity \( k \), then \( (i,j) \) must be constructed in a feasible solution. The feasible solution \( y \) and \( f \) then provides an upper bound to the primal problem by Eq.(1a), denoted by \( Z_{UB} \). Denote by \( Z_{UB}^{min} \) and \( Z_{LB}^{max} \) the best upper and the best lower bounds that have been obtained by the algorithm, respectively. The duality gap is then specified by \( GAP = Z_{UB}^{min} - Z_{LB} \). If \( GAP \) is more than the pre-defined threshold value (stop criterion; in the numerical example in Section
5 we use the threshold value 1), the algorithm then generates a new Lagrange multiplier vector $\lambda^{n+1}$ using the sub-gradient method, where $n$ denotes the iteration number. This process repeats until the duality gap $GAP$ is smaller than the pre-defined threshold value, at which point the stop criterion is satisfied. We apply the heuristic method proposed by Held and Karp (1971) to update $\lambda^{n+1}$:

$$\lambda^{n+1} = \max \{ \lambda^n + \theta^n \frac{Z_{UB}^{min} - LD(\lambda^n)}{\|s(\lambda^n)\|^2} s(\lambda^n), 0 \}$$

(9)

where $LD(\lambda^n)$ denotes an optimal solution to $LR$ with a given $\lambda^n$, $\theta^n$ is a step-size length (a scalar) with $0 \leq \theta^n \leq 2$, and $s(\lambda^n)$ defines the sub-gradient direction specified by the relaxed equation (2c), which is:

$$s(\lambda^n) = f_{ik}^j - y_{ij} \quad \forall (i,j) \in A, \forall k \in K$$

(10)

For the iterative technique used to determine successive values of $\lambda$, the choice of step size $\theta^n$ strongly affects the convergence of $LD$. The classical step-size rule satisfies $\theta^n \geq 0$, $\sum_{n=0}^{\infty} \theta^n = \infty$, and $\lim_{n \to \infty} \theta^n = 0$, for which the convergence proof is known.

The general choices of the step size have $\alpha/n$ or $\alpha/\sqrt{n}$, where $\alpha$ is a constant positive scalar. If the step size diminishes quickly, the convergence of the algorithm becomes slow. Hence, a step size that does not diminish quickly is desirable. In the numerical example in Section 5, we choose the step size $\theta = 5/\sqrt{n}$.

We generate a sequence $\lambda^0, \lambda^1, ..., \lambda^n$ and compute $Z_{LB}, Z_{UB}$ and $GAP$. The procedure is repeated until $GAP$ converges to the stop criterion. The algorithm is then terminated with the best feasible solution. The sub-gradient algorithm solving the $LD$ is presented as follows:

**Algorithm: Solving the Lagrangian dual problem (LD) for the PRT GN design**

1. **Step 1:** Set $n := 0$. Initialize a non-negative dual vector $\lambda^n$.
2. **Step 2:** Solve $LR(1)$ and $LR(2)$ with parameter $\lambda^n$. Compute $Z_{LB}$ using Eq.(5).
3. **Step 3:** Based on the solution of $LR(2)$, construct feasible solution of $y$; compute $Z_{UB}$ using Eq.(1a).
4. **Step 4:** Calculate duality gap $GAP$. If $GAP$ is small enough (smaller than a pre-defined threshold value), stop; otherwise go to Step 5.
5. **Step 5:** $n := n + 1$. Update $\lambda^n$ using Eqs. (9) – (10), and go to Step 2.

4.3. Generating cuts to tighten the Lagrangian lower bound

This section discusses three cuts to tighten the Lagrangian lower bound. Let $A_0$ denote the subset of arcs that has been identified as not to be constructed. Let $A_1$ denote the subset of arcs that has been identified as to be constructed.
(1) Fix \( y_{ij} = 0 \)

At each Lagrangian iteration when \( LR(1) \) is solved for, let \( c''' = \max_{(i,j) \in A}\{c'_{ij} | y_{ij} = 1\} \) and \( c''' = \min\{c'_{ij} | y_{ij} = 0\} \). Recall that we scan arcs in the increasing order of \( c'_{ij} \); we then identify the first arc \((v,w)\) such that \( c'_{vw} - \max\{0,c''\} \geq Z_{UB}^{min} - Z_{LB} \). The set of arcs whose order is more than that of \((v,w)\), i.e., \( \{(i,j) | r_{ij} \geq r_{vw}\} \), can be fixed by \( y_{ij} = 0 \) and added into \( A_0 \).

This can be explained as follows. Suppose \( y_{vw} = 1 \), then the objective function of \( LR(1) \) will increase. Consider the following two possibilities. (1) Suppose \( c'' > 0 \); it implies \( M_1 \) is binding, and the objective function value will increase by \( c'_{vw} - c'' \). (2) Suppose \( c'' \leq 0 \); it implies \( M_1 \) is not binding, and the objective function will increase by \( c'_{vw} \). In both cases the new objective function will increase by \( c'_{vw} - \max\{0,c''\} \). Suppose \( y_{vw} = 1 \), then the new Lagrangian lower bound equals \( Z_{LB} = Z_{LB} + c'_{vw} - \max\{0,c''\} \geq Z_{UB}^{min} \), which violates the Lagrangian duality theory and thus is not possible. The same result also applies to the set \( \{(i,j) | r_{ij} \geq r_{vw}\} \), which can be added to \( A_0 \) and be eliminated from the problem. \( M_2 \) can be tightened by \( M_2 := \min\{M_2, |A| - |A_0|\} \).

(2) Fix \( y_{ij} = 1 \)

At each Lagrangian iteration when \( LR(2) \) is solved for, we restrict \( y_{ij} = 0 \) one by one for each \( (i,j) \in A - A_0 - A_1 \) (the subset of arcs that has not been fixed). Meanwhile, \( y_{ij} = 1 \) in the solved solution \( y \). Let \((v,w)\) be such an arc. We then calculate the optimal Lagrangian lower bound subject to \( y_{vw} = 0 \) as follows. For \( LR(1) \), the penalty \( e_1 \) is subject to two possible conditions. If \( \sum_{(i,j) \in A} y_{ij} > M_1 \), there is \( e_1 = -c'_{vw} \); otherwise \( \sum_{(i,j) \in A} y_{ij} = M_1 \), there is \( e_1 = -c'_{vw} + c'''' \). To compute the penalty for \( LR(2) \), denoted by \( e_2 \), we restrict \( u_{vw} = 0 \) and solve the restricted \( LR(2) \) by LP. Let \( Z_R \) be the corresponding objective function value of the restricted \( LR(2) \), and \( Z_2 \) be the objective function value of the non-restricted \( LR(2) \). The penalty of \( LR(2) \) equals \( Z_R - Z_2 \). The addition of penalties imposed on both \( LR(1) \) and \( LR(2) \) is then compared with \( Z_{UB}^{min} - Z_{LB} \). If \( e_1 + e_2 > Z_{UB}^{min} - Z_{LB} \), we can fix \( y_{vw} = 1 \) and add \((v,w)\) into \( A_1 \). The logic is that suppose \( y_{vw} = 0 \); the new Lagrangian lower bound would be more than \( Z_{UB}^{min} \) and thus is impossible. \( M_1 \) can be tightened by \( M_1 = \max\{M_1, |A_1|\} \).

(3) Penalties on number of arcs

At each Lagrangian iteration when \( LR(1) \) is solved for, we solve the restricted problem by fixing \( \sum_{(i,j)} y_{ij} = M_1 \) and \( \sum_{(i,j)} y_{ij} = M_2 \), respectively. If the Lagrangian lower bound is more than the best upper bound \( Z_{UB}^{min} \), it implies the restricted problem is infeasible; therefore we then can tighten \( M_1 \) or \( M_2 \) by 1.
5. Numerical example

In this section, a real-sized example is analyzed to verify the PRT GN methodology and examine the algorithmic performance of the proposed method. The Lagrangian relaxation algorithm was coded in IBM ILOG CPLEX Optimization Studio 12.5.1 interfaced with MATLAB 2010b in Windows 7. Problem $LR(1)$ was solved by inspection; $LR(2)$ was a LP and solved using CPLEX. We solve the PRT GN problem on a PC equipped with a 2.66-GHz Intel(R) Xeon(R) E5640 CPU with 24 GB of memory.

The study network is based in the downtown area of Minneapolis, and is multi-modal as illustrated in Fig. 3. It contains 28 nodes and 43 arcs (two-way), with a total length of 140 miles. It consists of 3 light rail stations, 13 PRT stations and 6 Steiner points. The size of the example is relatively large compared to most PRT designs in current practice. To build the network we suppose that each mile of guideway costs $10M. Characteristics of the network are tabulated in Table 1.

(Place Table 1 here)

(Place Figure 3 here)

The hourly demand of the scenario is hypothetical. There are three light rail stations and the PRT network is connected with light rails. We assume that the demand between a light rail station and a PRT station is relatively large, and the demand between a pair of PRT stations is relatively small. Such a demand pattern is consistent with TOD deployment. The total demand in the example is 3,790 passengers during the peak hour. The ratio of the peak hour demand and 24-hour demand (K factor) is assumed to be 0.2.

The objective function specifies $\alpha = 0.5$ and $\beta = 0.5$. It implies that we place equal weight on the construction and user costs. To compute the user costs we use the value of time to convert the travel time to dollars, to make it comparable to the construction costs. In the evaluation of the life-cycle travel time cost for the PRT, we use 10 years to calculate the life cycle.

The proposed Lagrangian relaxation method can solve the two problems, with and without Steiner points, to a duality gap of less than 1. The problems are deemed to be solved to optimality with such small duality gaps. The solution statistics are compared in Table 2, and the solved PRT guideways are plotted in Fig. 4. Without the Steiner points, the total guideway length is 42.7 miles. As the construction cost is $10 million per mile per lane, and the construction is two lanes in each direction, the total construction cost is $854.2 million. With the Steiner points, the total guideway length is 34.4 miles, and the total construction cost is $687.4 million. Hence, when the Steiner points are considered, the construction costs reduce by about 20%. Without the Steiner points, the total user cost is $187.6 million (life cycle is 10 years), and with the Steiner points, the total user cost is reduced to $177.4 million. The improvement is about 5%. Hence, the introduction of the
Steiner points not only reduces the length (and investment) of the guideway, but also reduces the user travel time significantly. Fig. 4 illustrates that the introduction of the Steiner points can also lead to a significant difference in the GN topology.

(Place Table 2 here)

(Place Figure 4 here)

In terms of the computational performance, the scenario with the Steiner points takes up to 28 minutes to solve to optimality (duality gap less than 1). It runs up to 2,970 iterations. Note that in the first 120 iterations (around 100 seconds), the duality gap reduces from 173 to 12 rapidly. Then, there is a long tailing effect, which takes the majority of the time to reduce the duality gap from 12 to 1. It indicates that the algorithm can quickly find a relatively good feasible solution, and a majority of the convergence is achieved in the first 100 seconds. The scenario without the Steiner points runs much faster. It takes about 30 seconds to reach optimality using 83 iterations. It indicates that the GN design with Steiner points is harder to solve than the one without Steiner points. The proposed Lagrangian relaxation algorithm can solve both scenarios to optimality within a reasonable amount of time. The convergence performance of the algorithm with and without the Steiner points is plotted in Figure 5.

In general, the computational time is commensurate with the network size. Since the study network is relatively large compared to most existing PRT networks in practice, the computational time is relatively higher. For smaller networks, the computational time could be less.

(Place Figure 5 here)

6. Concluding remarks

This paper presents a model and solution algorithm for the GN design for the PRT supporting transit-oriented development. The GN is designed to balance two objectives, the life-time construction costs and the total user travel costs. We show that the introduction of a set of optional points, known as Steiner points, can reduce the GN length as well as the construction cost substantially. The resultant GN design problem is then formulated as a combined Steiner problem and assignment problem. A Lagrangian relaxation algorithm is proposed to decompose the Lagrangian dual problem into two subproblems, both of which are trivial to solve. A real-sized numerical example is used to demonstrate the computational performance and validate that a set of Steiner points can reduce both the construction and user travel costs significantly.

The proposed model is for the purpose of offline planning, or what-if scenario analysis, to assist PRT GN design supporting transit-oriented development. On a large-scale
network, the Lagrangian relaxation algorithm may take significantly more time to solve to a desired duality gap. This is a theoretical feature of the algorithm because the complexity of the formulated model is NP-hard. It is again pertinent to note here that the example network considered in this study is larger than PRTs currently in practice. The modest size of PRTs currently is because of the focused role of PRTs as an automated people mover in specific contexts. Hence, while there is always a tradeoff between computational performance and solution quality, we show that the proposed Lagrangian relaxation algorithm can solve a real-size PRT example to optimality within a reasonable amount of time.

Acknowledgements

This research was supported by the NEXTRANS Center, the USDOT Region V Regional University Transportation Center. The authors are solely responsible for the contents of the paper.

References


Juster, R. & Schonfeld, P. Year. An assessment of trip times in automated guideway transit systems. In: *Presented at the 2014 Annual Meeting of the Transportation Research Board,* 2013 Washington, DC.


Figure 1. An example of TOD development in Maryland (Source: (Central Maryland Transportation Alliance, 2009))
Figure 2. Steiner points to reduce the network length
Figure 3. Network in the numerical example
Two-way guideways
Light rail

Figure 4(a) Without Steiner points
Figure 4(b) With Steiner points

Figure 4. Guideway networks solved with and without Steiner points
Figure 5. Convergence performance of the Lagrangian relaxation algorithm
Table 1. Characteristics of the network.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guideway cost</td>
<td>$10 million/mile</td>
</tr>
<tr>
<td>Operating speed</td>
<td>30 mile/hr</td>
</tr>
<tr>
<td>Capacity</td>
<td>1,000 passenger cars</td>
</tr>
<tr>
<td>Value of time</td>
<td>$20 passenger/hr</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.5</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.5</td>
</tr>
<tr>
<td>K factor</td>
<td>0.2</td>
</tr>
<tr>
<td>Life cycle</td>
<td>10 years</td>
</tr>
</tbody>
</table>
Table 2. Solution characteristics.

<table>
<thead>
<tr>
<th></th>
<th>With Steiner points</th>
<th>Without Steiner Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guideway length (miles)</td>
<td>34.4</td>
<td>42.7</td>
</tr>
<tr>
<td>Construction costs (million $)</td>
<td>687.4</td>
<td>854.2</td>
</tr>
<tr>
<td>User costs (10 years, million $)</td>
<td>177.4</td>
<td>187.6</td>
</tr>
<tr>
<td>UB</td>
<td>522.1</td>
<td>600.4</td>
</tr>
<tr>
<td>Lagrangian LB</td>
<td>521.1</td>
<td>599.4</td>
</tr>
<tr>
<td>Duality gap</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Computational time (secs)</td>
<td>1689.1</td>
<td>33.5</td>
</tr>
</tbody>
</table>
List of Figures
Figure 1. An example of TOD development in Maryland (Source: (Central Maryland Transportation Alliance, 2009)) ................................................................. 18
Figure 2. Steiner points to reduce the network length .................................................................................. 19
Figure 3. Network in the numerical example ................................................................................................. 20
Figure 4. Guideway networks solved with and without Steiner points ......................................................... 22
Figure 5. Convergence performance of the Lagrangian relaxation algorithm ............................................. 23
List of Tables

Table 1. Characteristics of the network.......................................................... 24
Table 2. Solution characteristics. ...................................................................... 25
List of Keywords:

network design,

Steiner problem,

Lagrangian relaxation,

personal rapid transit,

transit-oriented development.