Integrated Modeling of High Performance Passenger and Freight Train Planning On Shared-Use Corridors in the US

February 2015

PROJECT NO.
UIC 2117-9060-02-C

PREPARED BY
The Urban Transportation Center at the University of Illinois at Chicago
INTEGRATED MODELING OF HIGH PERFORMANCE PASSENGER AND FREIGHT TRAIN PLANNING ON SHARED-USE CORRIDORS IN THE US

February 2015

This report was produced with funding from:

National Center for Transit Research (NCTR), a US DOT-OST National University Transportation Center

Metropolitan Transportation Support Initiative (METS1)
INTEGRATED MODELING OF HIGH PERFORMANCE PASSENGER AND FREIGHT TRAIN PLANNING ON SHARED-USE CORRIDORS IN THE US

Ahmadreza Talebian, Bo Zou

Department of Civil and Materials Engineering, University of Illinois at Chicago, United States

Abstract: This paper studies strategic level train planning for high performance passenger and freight train operations on shared-use corridors in the US. We develop a hypergraph-based, two-level approach to sequentially minimize passenger and freight costs while scheduling train services. Passenger schedule delay and freight lost demand are explicitly modeled. We explore different solution strategies and conclude that a problem-tailored linearized reformulation yields superior computational performance. Using realistic parameter values, our numerical experiments show that passenger cost due to schedule delay is comparable to in-vehicle travel time cost and rail fare. In most cases, marginal freight cost increase from scheduling more passenger trains is higher than marginal reduction in passenger schedule delay cost. The heterogeneity of train speed reduces the number of freight trains that can run on a corridor. Greater tolerance for delays could reduce lost demand and overall cost on the freight side. The approach developed in the paper could be applied to other scenarios with different parameter values.

Key words: shared-use rail corridor, strategic planning, hypergraphs, integer programming, passenger schedule delay, lost freight demand

1 Introduction

Passenger rail has been resurging in the US. Amtrak, the primary intercity rail service provider, has witnessed 51.1% ridership growth between 2000 and 2013, from 20.9 to 31.6 million passengers (Amtrak, 2013). To sustain this trend and promote sustainability and multimodality for inter-city travel, several states have been pursuing high performance rail systems. Among many, California has started building a $68 billion brand-new, dedicated High Speed Rail (HSR) line (California HSR, 2012). The Midwest region takes a more conservative approach by taking advantage of using existing rail infrastructure by both passenger and freight trains (Peterman et al., 2009). One prominent example is the Chicago – St. Louis corridor, where the existing single-track line owned by Union Pacific railroad is being upgraded to accommodate future passenger services running at up to 110 mph. Once track upgrade is completed in 2017, the travel time between Chicago and St. Louis will be reduced by one hour (Illinois HSR, 2014). By utilizing existing infrastructure, the project cost is much lower than California HSR, only a few billion dollars for the initial phase (Illinois HSR, 2014). Given the economic appeal of high performance passenger rail on shared-use corridors and the fact that such services have not been put in place, it is important to understand, from the strategic planning perspective, the interactions between passenger and freight services.

Strategic planning on shared-use corridors has different meanings for passenger and freight trains. On the passenger side, strategic planning pertains to determining preliminary, non-minute-by-minute train schedules given passenger demand and the number of trains. It is part of the scheduling stage of a typical six-stage rail planning process (Ghoseiri et al., 2004). On the freight side, strategic level planning refers to modeling train operations taking account of overall freight demand and the number of programmed services, without going into detailed operating characteristics. This is particular relevant in the US, where freight train departures are a function of demand: a train simply departs once it receives sufficient load (Cordeau et al., 1998).
Because Amtrak is a publically funded entity, we assume in this study that Amtrak cares about passenger benefits. One important measure related to passenger benefits is passenger schedule delay. For a passenger, schedule delay is the time difference between one’s preferred and actual departure (Hendrickson and Kocur, 1981), which characterizes the inconvenience of transportation service schedule to accommodate one’s desired activities. In passenger rail service planning, very limited attention has been paid to schedule delay. For non-rail modes, Kanafani (1983) shows that in case of air passenger transportation, cost due to schedule delay is comparable to other cost components in short-haul markets. Given that train services are typically less frequent than flights in the US, we expect schedule delay plays an even more important role in planning passenger rail services.

Quantifying passenger schedule delay requires knowledge about both the departure time of trains and the distribution of passenger preferred departure time (PDT). Modern travel survey methods, travel demand forecasting techniques, and automated passenger counting systems have made possible constructing distribution profiles of passenger PDTs, in which PDTs are discretized into relatively coarse time intervals (e.g., 15-min intervals as in Cascetta and Coppola (2012)). For consistency, it is sensible to consider train schedules with similar time resolution.

The objective of the paper is to investigate strategic level rail planning on shared-use corridors with the presence of high performance passenger trains. We recognize that, by the US Federal law (110 Congress, 2008; Harrod, 2009; Wilner, 2013), passenger trains are given access priority over freight operations on shared-use corridors. We employ a hypergraph-based, two-level nonlinear integer programming model to sequentially determine passenger and freight train schedules. This depicts the scenario in which the federal law is effectively applied. Passenger schedule delay and lost freight demand due to infrastructure capacity constraints are explicitly incorporated in the modeling process. By exploring different solution strategies, we conclude that a problem-tailored linearized reformulation of the original model yields superior computational performance. Using realistic parameter values, our numerical experiments show that passenger cost due to schedule delay is comparable to in-vehicle travel time cost and rail fare. We find that in most cases marginal freight cost increase from scheduling more passenger trains is higher than marginal reduction in passenger schedule delay cost. The heterogeneity of train speed reduces the number of freight trains that can run on a shared-use corridor. Greater tolerance for delays of the freight operator could reduce lost demand and overall freight side cost. The approach developed in the paper could be applied to other scenarios with different parameter values.

We begin with a review of train scheduling literature in Section 2. The mathematical formulation of our problem is presented in Section 3. We discuss on the solution strategies in Section 4. Section 5 performs numerical analyses, on both a sample problem and a simplified case study for the Chicago-St Louis HSR corridor. Sensitivity analysis on speed heterogeneity and freight train delay tolerance are also conducted. Section 6 summarizes major findings and offers directions for future research.

2 Literature review and research contribution

The research on train scheduling dates back to as early as Frank (1966). Models developed since then fall into three categories: analytical, simulation, and discrete optimization approaches (Abril et al., 2008). The analytical approach uses simple models to estimate rail line capacity, train delay, and cycle times through probabilistic or deterministic analysis of train dispatching patterns (e.g., Chen and Harker, 1990; Hallowell and Harker, 1996; Flier et al., 2009). However, the simplicity of this approach limits its capability in dealing with complex real world situations.

Simulation is the dominant method in practice. Commercial software such as Rail Traffic Controller (RTC) (Wilson, 2012) and Módulo Optimizador de Mallas (MOM) (Barber et al., 2006) incorporate a range of parameters, including train types, equipment types, terrain and track conditions, train speed, acceleration and deceleration, and traffic signals, to reflect train dispatching and operation in the real world. In general, the simulation approach does not seek to optimize train schedules, unless combined with optimization techniques (e.g., Jovanović and Harker, 1991). In addition, passenger cost is not considered while simulating train schedules.

The discrete optimization approach uses mathematical programming to identify train schedules that correspond to user-defined system optimum in fairly complex situations. Under this approach, train scheduling is most commonly modeled with discrete time networks, multi-commodity flows, and constrained resources. Since the first study by Amit and Goldfarb (1971), the literature has grown substantially. We review below only some of the more recent papers that are relevant to our study. Carey and Lockwood (1995) formulate a 0-1
mixed integer program for train dispatching on a single uni-directional line where overtaking is allowed. The objective is to minimize total cost which is the sum of cost of deviating from preferred departure and arrival; cost of travel times on links; and cost of dwell times at stations. To solve large-scale problems, the authors propose an iterative decomposition approach that is analogous to manual search strategies which human train planners found very effective in practice. The research is extended to more general networks with choice of lines, platforms, and routes (Carey, 1994a), and two-way track (1994b). Brännlund et al. (1998) present a binary linear program for scheduling passenger and freight trains on a single-track corridor. The objective is to maximize profits of all trains subject to track capacity constraints. Using Lagrangian relaxation, the original problem is decomposed into independent shortest path subproblems, one for each train. Caprara et al. (2002; 2006) seek train timetables with least deviation from the ideal ones, given track capacity constraints. Binary linear programs are formulated to maximize the sum of profits across arcs, which is equivalent to minimizing schedule deviation. Cacchiani et al. (2010) build on Caprara et al.'s models to further allow for bi-directional traffic, rerouting, and arbitrary network topology. A new arc coupling formulation is proposed by Borndörfer and Schlechte (2007), Borndörfer et al. (2010), and Schlechte (2012), in which the rail network is modeled as an expansion of a track without occupation conflicts. To solve large scale problems, branch-and-price heuristics combined with column generation and bundle methods are introduced. Harrod (2009; 2011) develops a hypergraph-based approach which explicitly handles path conflicts while trains are transitioning between blocks. Such conflicts are often omitted in conventional time-block occupancy-based formulations.

The concept of job-shop scheduling has also been applied to train scheduling research. Train trips are considered as a set of jobs which are scheduled on resources (tracks). In Oliveira and Smith (2000), total train departure delays are minimized with respect to ideal timetables, subject to requirements for train meeting, minimum headway, and potential track blocking. When a subgroup of trains has scheduling priority, priority constraints are further introduced (Liu and Kozan, 2011). The objective of minimizing makespan in Liu and Kozan (2011) translates into minimizing total length of train schedules. Zhou and Zhong (2005) consider multi-mode flow-shop scheduling for high- and medium-speed trains. The scheduling has two objectives: minimizing interdeparture time variation for high speed trains and total travel time for all trains. A beam search algorithm is employed to generate non-dominated train schedules.

Despite the large body of train scheduling literature, one important aspect that has not received sufficient attention is passenger side effects. Only a few studies have looked into the trade-off between passenger travel time and train operation performance. Ghoseiri et al. (2004) construct a Pareto frontier and identify optimal schedules that minimize the weighted sum of train fuel consumption and passenger travel time. Carbon emission cost is further considered in Li et al. (2013) using a fuzzy multi-objective optimization algorithm. A goal programming approach is employed by Yang et al. (2009) to minimize the fuzzy total of passenger in-vehicle time and total train delay at stations. A branch-and-bound algorithm with fuzzy simulation is used to obtain optimal solutions.

None of the existing studies deal explicitly with schedule delay of rail passengers, although passenger waiting time as a related issue has been considered. Ceder (1991) presents a model which aims at reducing passenger waiting while maintaining the minimum number of trains required. Nachtigall (1996) minimizes the weighted sum of passenger waiting times by setting a boundary for each train running and stopping. Canca et al. (2012) assumes that travel demand between two stations increases to the extent that a new shuttle service is required to serve the new demand. They try to insert the optimal number of shuttle trains which skip stations that have low demand, so that total passenger waiting is minimized. Canca et al. (2011; 2014a) formulate non-linear integer programs to minimize passenger average waiting time. Canca et al. (2014b) extend Canca et al. (2014a) to incorporate demand elasticity with respect to the number of trains. The problem is formulated as minimizing the probability of rail passenger loss, using logit-based and sigmoid function-based models. Finally, linearized formulations and a fast adaptive large neighborhood search metaheuristic are proposed by Barrena et al. (2014a, b) to minimize passenger waiting at stations. The last three studies focus exclusively on passenger traffic in a one-way, single train type environment.

3 The model

In this section we propose a hypergraph-based, two-level sequential approach to schedule passenger and freight train operations on a shared-use corridor. We first introduce hypergraphs to characterize train movements and then formulate the train scheduling models. The sets, parameters, and decision variables used in the model are documented in Table 1.
### Table 1: Notations used in the models

<table>
<thead>
<tr>
<th>Type</th>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary decision</td>
<td>$x^r_{i,j,u,v}$</td>
<td>Occupancy arc denoting if (sub)train $r$ enters into block $i$ at $u$, occupies block $i$ in time interval $[u,v)$, and exits into block $j$ at time $v$</td>
</tr>
<tr>
<td>variables</td>
<td>$y^r_{i,t}$</td>
<td>Artificial linking arc denoting if the arrival of subtrain $r$ at its destination at time $t$ is linked to the departure of its continuation subtrain $\hat{r}$ at time $\hat{t}$</td>
</tr>
<tr>
<td>Sink node</td>
<td>$\epsilon^r$</td>
<td>Artificial sink designating that (sub)train $r$ is off the network</td>
</tr>
<tr>
<td>Parameters</td>
<td>$w^w$</td>
<td>Origin block of station pair $w$</td>
</tr>
<tr>
<td></td>
<td>$d^w$</td>
<td>Destination block of station pair $w$</td>
</tr>
<tr>
<td></td>
<td>$EADT^r$</td>
<td>Earliest allowed departure time (EADT) from origin of (sub)train $r$</td>
</tr>
<tr>
<td></td>
<td>$LAAT^r$</td>
<td>Latest allowed arrival time (LAAT) at destination of (sub)train $r$</td>
</tr>
<tr>
<td></td>
<td>$l^r_{\text{max}}$</td>
<td>Maximum allowable layover time at a station for passenger subtrain $r$</td>
</tr>
<tr>
<td></td>
<td>$l^r_{\text{min}}$</td>
<td>Minimum allowable layover time at a station for passenger subtrain $r$</td>
</tr>
<tr>
<td></td>
<td>$c^r_i$</td>
<td>Cost of layover for passenger subtrain $r$ measured in $$/unit time</td>
</tr>
<tr>
<td></td>
<td>$c^r_p$</td>
<td>Lost demand cost for freight train $r$ measured in $$/train</td>
</tr>
<tr>
<td></td>
<td>$c^r_\ell$</td>
<td>Freight value of time for freight train $r$ measured in $$/unit time</td>
</tr>
<tr>
<td></td>
<td>$c^r_w$</td>
<td>Train operating cost of stopping status plus freight value of time for freight train $r$ measured in $$/train</td>
</tr>
<tr>
<td></td>
<td>$b^i_t$</td>
<td>Capacity (number of trains) of block $i$ at time $t$</td>
</tr>
<tr>
<td></td>
<td>$k^i_t$</td>
<td>Capacity (number of trains) of cell $i$ at time $t$</td>
</tr>
<tr>
<td></td>
<td>$\xi$</td>
<td>Leading transition time margin</td>
</tr>
<tr>
<td></td>
<td>$\delta$</td>
<td>Lagging transition time margin</td>
</tr>
<tr>
<td></td>
<td>$h^r$</td>
<td>Minimum gap between (sub)train $r$ and following (sub)trains</td>
</tr>
<tr>
<td></td>
<td>$c^r_\ell_d$</td>
<td>Value of schedule delay time for travelers whose PDT is earlier than the actual departure, measured in $$/time unit</td>
</tr>
<tr>
<td></td>
<td>$c^r_\ell_r$</td>
<td>Value of schedule delay time for travelers whose PDT is later than the actual departure, measured in $$/time unit</td>
</tr>
<tr>
<td></td>
<td>$q^w_w$</td>
<td>Total number of passengers leaving the origin of station pair $w$ towards the destination of station pair $w$ (consisting of those with the destination of station pair $w$ as their true destination station and those whose final destination station is in the same direction but beyond the destination of station pair $w$), and desire to leave between $t=m-1$ and $t=m$.</td>
</tr>
<tr>
<td></td>
<td>$r^w_n$</td>
<td>$n$th subtrain running between station pair $w$</td>
</tr>
<tr>
<td>Sets</td>
<td>$T$</td>
<td>The discrete-time horizon, ordered with starting value 1</td>
</tr>
<tr>
<td></td>
<td>$R$</td>
<td>The set of all passenger subtrains and freight trains</td>
</tr>
<tr>
<td></td>
<td>$R^p$</td>
<td>The subset of passenger subtrains, $R^p \subset R$</td>
</tr>
<tr>
<td></td>
<td>$R^{p,N}$</td>
<td>The set of passenger subtrains traveling in the direction with increasing track block index</td>
</tr>
<tr>
<td></td>
<td>$R^{p,S}$</td>
<td>The set of passenger subtrains traveling in the direction with decreasing track block index. $R^{p,N} \cup R^{p,S} = R^p$</td>
</tr>
<tr>
<td></td>
<td>$R^f$</td>
<td>The subset of freight trains. $R^p \cup R^f = R$</td>
</tr>
<tr>
<td></td>
<td>$R^{f,N}$</td>
<td>The set of freight trains traveling in the direction with increasing track block index</td>
</tr>
<tr>
<td></td>
<td>$R^{f,S}$</td>
<td>The set of freight trains traveling in the direction with decreasing track block index. $R^{f,N} \cup R^{f,S} = R^f$</td>
</tr>
<tr>
<td></td>
<td>$B$</td>
<td>The set of track blocks</td>
</tr>
<tr>
<td></td>
<td>$Z^p$</td>
<td>The set of linked passenger subtrains ${(r,\hat{r})}$, where $r$ is a terminating subtrain and $\hat{r}$ is an originating subtrain at the same location. Both subtrains refer to the same physical train.</td>
</tr>
<tr>
<td></td>
<td>$\Psi^p_r$</td>
<td>The set of feasible path arcs $(i,j,u,v)$ for passenger subtrain $r$ supplied from preprocessing</td>
</tr>
<tr>
<td></td>
<td>$\Psi^f_r$</td>
<td>The set of feasible path arcs $(i,j,u,v)$ for freight train $r$ supplied from preprocessing</td>
</tr>
<tr>
<td></td>
<td>$T$</td>
<td>The set of network cells</td>
</tr>
<tr>
<td></td>
<td>$L^p_{r,t}$</td>
<td>The set of valid pairs of arrival time of passenger subtrain $r$ and departure time of its continuation subtrain $\hat{r}$: $(t,\hat{t}) \in T$</td>
</tr>
<tr>
<td>Type</td>
<td>Component</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>$T$</td>
<td></td>
<td>Set of time periods considered for train scheduling</td>
</tr>
<tr>
<td>$W$</td>
<td></td>
<td>Set of station pairs</td>
</tr>
<tr>
<td>$C_{u,\mathcal{L}}$</td>
<td>Total schedule delay cost of passengers who prefer to depart between $t=1$ and $u$ (departure time of subtrain $r^w_n$) and end up boarding subtrain $r^w_1$</td>
<td></td>
</tr>
<tr>
<td>$C_{u,\mathcal{R}}$</td>
<td>Total schedule delay cost of passengers who prefer to depart between $u$ (departure time of the subtrain $r^w_n$) and $c_d\mathcal{L}_u+c_d\mathcal{R}_u$, where $u'$ is a given departure time of subtrain $r^w_n-1$, and end up boarding subtrain $r^w_n$</td>
<td></td>
</tr>
<tr>
<td>$C_{u,\mathcal{L}}$</td>
<td>Total schedule delay cost of passengers who prefer to depart before $u$ (departure time of subtrain $r^w_n$) and end up boarding subtrain $r^w_n$ (please refer to text to see difference between $C_{u,\mathcal{L}}$ and $C_{u,\mathcal{R}}$)</td>
<td></td>
</tr>
<tr>
<td>$C_{u,\mathcal{R}}$</td>
<td>Total schedule delay cost of passengers who prefer to depart after $u$ (departure time of subtrain $r^w_n$) and end up boarding subtrain $r^w_n$ (please refer to text to see difference between $C_{u,\mathcal{L}}$ and $C_{u,\mathcal{R}}$)</td>
<td></td>
</tr>
<tr>
<td>$C_{u,\mathcal{L}}$</td>
<td>Total schedule delay cost of passengers who prefer to depart between $u$ (departure time of subtrain $r^w_n$) and $t=T$, and end up boarding subtrain $r^w_n$</td>
<td></td>
</tr>
</tbody>
</table>

### 3.1 Hypergraphs for characterizing train movements

A hypergraph is a generalized graph which allows an edge (or arc) to connect multiple nodes. In this paper, we follow Harrod (2011) and use directed hypergraphs to model train moves on rail tracks. An example of this type of hypergraph is shown in Figure 1. These hypergraphs can contain two types of nodes: block occupancy nodes (solid points in Figure 1) and cells (hollow points). Each block occupancy node is labeled by its block (track) number and occupation time: $(i, t), i \in B, t \in T$. Each cell, capturing train transition between neighboring blocks, is labeled by its lower-left node. For instance, the outlined cell in Figure 1 will be labeled $(3, t)$. A simple hyperarc covering two block occupancy nodes and one cell is shown by the circled arrow in Figure 1.

![Figure 1: An illustration of hypergraphs (adapted from Harrod (2011))](image)

We describe train movements using hyperarcs. Specifically, we introduce binary variables $x_{i,j,u,v}^r$ to indicate whether train $r$ occupies block $i$ at time $u$ and exits into block $j$ at time $v$. For example, the circled hyperarc in Figure 1 can be represented by $x_{1,2,t,t+1}^r$. Obviously, if $i = j$ then there is no movement and the associated hyperarc will be horizontally positioned. No cell will be involved.

A chain of consecutive hyperarcs forms a train path. To illustrate, the grey path (for train $g$) in Figure 2 is characterized by hyperarcs $x_{5,4,t,t+1}^g, x_{4,3,t+1,t+2}^g, x_{3,3,t+2,t+3}^g, x_{3,3,t+3,t+4}^g, x_{3,2,t+4,t+5}^g$, and $x_{3,1,t+5,t+6}^g, x_{3,3,t+2,t+3}^g$ and
\(x^g_{3,3,t+3,t+4}\) denote train stopping at block 2; whereas \(x^g_{3,3,t+3,t+2}, x^g_{3,2,t+4,t+5}\) and \(x^g_{2,1,t+5,t+6}\) correspond to train moves. In this study we restrict \(i\) and \(j\) in \(x^r_{i,j,u,v}\) to be either identical (stopping) or two neighbored blocks (moving), i.e., \(j\) can only take three values \(i, i + 1,\) or \(i - 1\), the latter two associated with moving in opposite directions. Under this restriction, \(x^r_{i,j,u,v}\) then indicates whether train \(r\) occupies block \(i\) in time interval \([u, v)\) and exits into block \(j\) at time \(v\). We further assume that each stopping hyperarc takes one unit of time, i.e., if \(i = j\), then \(v = u + 1\) in \(x^r_{i,j,u,v}\). In contrast, a moving arc can take multiple time units depending on block length and train speed specified.\(^2\) With these assumptions, a given \(x^r_{i,j,u,v}\) uniquely identifies a hyperarc.

As described in detail by Harrod (2011), hyperarcs allow for explicit characterization of potential path conflicts while trains transition between two neighboring blocks, through the intermediate cell. This is done by specifying a transition window \([t + 1 - \epsilon, ..., t + 1 + \delta]\) for each cell, where \(\epsilon\) and \(\delta\) are leading and lagging time margins for transition. As we will show later in sub-Section 3.2, train paths are operationally feasible with respect to a cell when the capacity of the cell at \(t\) is not exceeded by the number of active hyperarcs covering the cell within the transition window \([t + 1 - \epsilon, ..., t + 1 + \delta]\).\(^3\)

![Figure 2: Two representations of train paths using hypergraphs](image)

To facilitate modeling passenger train movements, we further introduce subtrains. A physical passenger train can be decomposed into subtrains, each making a journey between two consecutive stations. In this study we assume that each physical passenger train running on a line stops at the same stations. We use \(r^w_n\) \((n = 1, ..., N; w \in W)\) to denote the \(n^{th}\) subtrain traveling on (directional) station pair \(w\). We use Figure 3 to illustrate each of the three physical trains traveling from station 1 to station 4 consists of three subtrains. \(r^w_1\) is the first subtrain of the first physical train traveling from station 1 to station 2 (station pair \(w_1\)). \(r^w_2\) is the second subtrain of the first physical train, which is linked to \(r^w_1\) and runs from station 2 to station 3 (station pair \(w_2\)). \(r^w_3\) is the first subtrain of the second physical train traveling on station pair \(w_1\). In the rest of the paper, superscript \(r\) in \(x^r_{i,j,u,v}\) may be replaced by \(r^w_n\) if \(r\) denotes a subtrain and information on \(w\) and \(n\) is needed. Note that a special case is when a train's journey involves only two stations. In this case, a physical train is identical to its only subtrain.

To complete the characterization of train movements, two types of artificial arcs – describing how a subtrain ends its journey and how a subtrain is linked to its subsequent one – need to be specified. The first type of arcs, termed sinking arcs, removes subtrain \(r^w_n\) off the line when it arrives at its destination station block.

\(^2\) Other types of hyperarcs may also be considered to represent train paths. For example, use a single hyperarc to denote a full train path (the black train in Figure 2). In this case, the hyperarc train path covers seven block occupancy nodes and four cells. Adopting higher dimension hyperarcs would require different preprocessing to generate feasible train paths.

\(^3\) Alternatively, one can employ a non-hypergraph-based approach, such as the one in Brännlund et al (1998) or simply letting trains occupy blocks for more time periods, to account for train path conflicts during transitions between blocks in an implicit way.
$d^w$, to an artificial sink $e^r$ (which can be viewed as an artificial block). Similar to the hyperarc notation, sinking arcs can be expressed using binary variables $x^{w}_{d^w,e^r,u,v}$. We assume that this removal takes one unit of time, i.e., if $x^{w}_{d^w,e^r,u,v} = 1$, then $v = u + 1$.

The second type of artificial arcs, termed linking arcs, links one subtrain to its very next one. Once a subtrain $r$ arrives at the artificial sink $e^r$ at $t$, we use a binary variable $y^{r,t}_{t+1}$ to denote if subtrain $\hat{r}$ (of the same physical train) will be its continuation and start the journey from the station where $r$ sank at time $\hat{t}$. $\hat{t} - t$ then denotes the layover time at the station (for passenger boarding and alighting) of the physical train. In this study the layover time of each subtrain $r$ at its destination station is bounded by pre-specified values $l^{r}_{\text{max}}$ and $l^{r}_{\text{min}}$.

We use Figure 4 to further demonstrate this. The physical train consists of two subtrains ($r$ and $\hat{r}$). $r$ starts its journey from station 1 (block 1) at $t$ and arrives at station 2 (block 3) at $t+2$. The first grey hyperarc is a sinking arc leading subtrain $r$ to artificial sink $e^r$. The second grey arc is a linking arc $(y^{r,F}_{t+3,t+4})$ linking subtrain $r$ to $\hat{r}$. $\hat{r}$ then departs at $t+4$ and finishes its journey at $t+7$.

3.2 Model formulation

Recall that by the US federal law passenger rail services are given access priority on shared-use corridors. To reflect this, we adopt a two-level sequential approach to perform strategic level train planning. At the upper level, we first identify passenger train schedules that minimize passenger side cost. Given passenger train
schedules, at the lower level we develop freight train schedules that minimize freight side cost. We provide a detailed model formulation in this subsection.

3.2.1 Upper Level: passenger train scheduling

Total passenger side cost consists of costs to: 1) passenger schedule delay; 2) passenger in-vehicle travel time; and 3) train operations. Ticket fare is internal transfer between passengers and Amtrak, and therefore is not counted. We consider schedules that permit two opposing passenger trains to pass without full stop (i.e., zero speed), called “flying meet” (Dure, 1999; Petersen and Taylor, 1987). This is done by specifying the set of feasible path arcs $Ψ^r$ for each subtrain $r$, in which no en-route stopping arc is generated. In this study we consider that sidings are long enough to safely accommodate such a meet. Assuming constant train speeds and flying meets, passenger in-vehicle travel time and train operating cost will be invariant to train schedules (though both assumptions of constant train speed and flying meets can be relaxed. See our discussion and further investigations in Appendices 1 and 2). The upper level problem then reduces to minimizing schedule delay cost of all rail travelers.

We calculate passenger schedule delay cost based on passenger Preferred Departure Time (PDT) profiles and train schedules. Each PDT provides total passenger demand originating from a station for a travel direction and its distribution over discrete time intervals in a day. The passengers consist of those with the next station as the final destination and those riding trains beyond that station. Figure 5 shows an exemplary passenger PDT with three train departures. Each bar indicates the number of travelers whose PDT falls within a time interval. For strategic level planning, we consider trains to have sufficient seat capacity. Each passenger selects a train that minimizes his/her schedule delay. If the passenger’s PDT does not coincide exactly with the departure time of a train, then some displacement of the actual departure time from his/her PDT will result, which is his/her schedule delay. If a traveler values equally the inconvenience caused by one unit of earlier and later displacement, then the closest train to his/her PDT will be chosen. It is possible that later displacement causes greater inconvenience than earlier displacement, especially when one wants to avoid delayed arrival. In this case, a larger schedule delay penalty will be assigned to later displacement. By performing this train selection process over all passengers, we will obtain the number of passengers boarding each train and total passenger schedule delay cost associated with each train. Passengers may alternatively care about Preferred Arrival Time (PAT). As we assume no train delays en route, the distribution of PAT for each origin-destination (OD) pair can be easily constructed using the distribution of PDT and in-vehicle travel time between each OD.

Note that passenger trains are not fully flexible in terms of when they depart and arrive. This is because the schedule of a train is constrained by the schedule of other trains. In this study an earliest allowed departure time ($EADT$) and a latest allowed arrival time ($LAAT$) are assigned to each subtrain. The $EADT$ of the first subtrain of the first physical train (which runs on station pair $w_1$), $EADT^{1w_1}$, is set equal to the beginning of the planning horizon $T$. $EADT^{1w_1}$ is then the beginning of $T$ plus unimpeded travel time of $r_1^{w_1}$ on $w_1$. We repeat the calculation of $EADT$'s for subsequent subtrains of the first physical train and also for all remaining physical trains assuming a minimum headway. $LAAT$ is calculated similarly: for each physical train, we set $LAAT$ of the last subtrain of the last physical train, $EADT^{Lw_L}$ (L denotes the last station pair), equal to the end of $T$, and then compute $LAAT$ backwards iteratively till the first subtrain of the same physical train using unimpeded travel time between station pairs. This is repeated over all other physical trains.
The first subtrain on station pair \( w \) (\( r_1^w \)) will attract all travelers from two groups: those preferring to depart before the departure of \( r_1^w \), and part of travelers whose PDT falls between the departures of \( r_1^w \) and \( r_2^w \). The first group of travelers will make later displacement; the second group will make earlier displacement. Recall that traveler inconveniences caused by earlier and later displacement can be different. We introduce two cost coefficients \( c_2^w \) and \( c_3^w \) to denote the respective unit cost (\$/time unit) for a traveler. For the first group of travelers, we compute passenger schedule delay cost for each time interval in the PDT profile that is earlier than the departure of \( r_1^w \), and then aggregate over all time intervals to obtain total passenger schedule delay cost for this group. Specifically, we multiply the number of passengers with PDT in the \( m \)-th time interval \( (q_m^w) \) by the schedule delay per passenger in the interval: \( u − \frac{m−1+m}{2} \), where \( u \) is the departure time of subtrain \( r_1^w \) and the midpoint in the \( m \)-th interval \( \frac{m−1+m}{2} \) is used as the reference time point for passengers’ PDT, and further by cost coefficient \( c_2^w \). We sum over all time intervals from the beginning of the day \((m=1)\) until \( u \) \((m=u)\). The resulting schedule delay cost for passengers whose PDT is earlier than the departure of \( r_1^w \) is expressed by (1.1).

For travelers in the second group, those choose \( r_1^w \) because doing so gives them smaller schedule delay. Let \( u' \) denote the departure of \( r_2^w \) \((u' > u)\). As the PDT of travelers gets farther away from \( u \) and closer to \( u' \), they are more likely to choose \( r_2^w \). There is a cutoff time point \( t_c \) between \( u \) and \( u' \), such that travelers with PDT at \( t_c \) will be indifferent to taking either subtrain. The indifference suggests that \( c_3^w * (t_c − u) = (u' − t_c) * c_2^w \), which yields \( t_c = \frac{c_2^w u + c_3^w u'}{c_2^w + c_3^w} \). Clearly, if one’s PDT is earlier than \( t_c \), \( r_1^w \) will be chosen. To calculate schedule delay cost of passengers in the second group, we similarly multiply the number of passengers with PDT in time interval \( m \) \((q_m^w)\) by the schedule delay per passenger in the interval \( \frac{m−1+m}{2} − u \) and by the cost coefficient \( c_3^w \). We sum over all time intervals from \( u + 1 \) to the cutoff point \( \left\lfloor \frac{c_2^w u + c_3^w u'}{c_2^w + c_3^w} \right\rfloor \) (we use floor function as an approximation). This gives the total schedule delay cost for passengers whose PDT is later than the departure of \( r_1^w \) and taking \( r_1^w \).

For the second to the second-to-last subtrain (i.e., \( n = 2,3, ..., N−1 \) for \( r_n^w \)), passenger schedule delay costs are calculated in a similar fashion. (1.2) shows schedule delay cost of passengers with PDT between the departures of \( r_n^w \) and \( r_{n+1}^w \), and end up taking \( r_n^w \) (i.e., earlier displacement). The calculation is exactly the same as described in the previous paragraph. (1.4) presents schedule delay cost of passengers with PDT between the departures of \( r_{n−1}^w \) and \( r_n^w \) \((n = 2, ..., N)\) and taking \( r_n^w \) (i.e., later displacement). For travelers preferring to depart after the departure of the last subtrain \( r_N^w \), schedule delay calculation follows the same logic as that for travelers preferring to depart before the departure of \( r_1^w \). The schedule delay cost is expressed by (1.6).

To simplify notation in the subsequent cost minimization formulation (see objective function (2.1)), we introduce two new expressions (1.3) and (1.5), which come from multiplying (1.2) and (1.4) by \( x_{o,w,f,u,u'}^{n+1} \) and \( x_{o,w,f,u,u'}^{n-1} \) respectively, and then summing over all possible \( u' \).
\[ C_{u,w}^{w} = \sum_{m=1}^{u} c_{d}^{w} q_{m}^{w} \left( u - \frac{m - 1 + m}{2} \right) \]  
\forall w \in W, \forall u \in \{ (o^{w}, j, u, v) \in \Psi^{p,r^{w}} \}  
(1.1)

\[ C_{u,w}^{w} = \sum_{m=1}^{u} c_{d}^{w} q_{m}^{w} \left( u - \frac{m - 1 + m}{2} \right) \]  
\forall w \in W, \forall u \in \{ (o^{w}, j, u, v) \in \Psi^{p,r^{w}} \}  
(1.2)

\[ C_{u,u'}^{w} = \sum_{u' | u < u'} C_{u,w}^{w} \times x_{o^{u'}, j, u', v'}^{w+1} \]  
\forall w \in W, \forall n = 1, 2, \ldots, N - 1, \forall u \in \{ (o^{w}, j, u, v) \in \Psi^{p,r^{w}} \}  
(1.3)

\[ C_{u,u'}^{w} = \sum_{u' | u < u'} C_{u,w}^{w} \times x_{o^{u'}, j, u', v'}^{w-1} \]  
\forall w \in W, \forall n = 2, 3, \ldots, N, \forall u \in \{ (o^{w}, j, u, v) \in \Psi^{p,r^{w}} \}  
(1.4)

\[ C_{u,u'}^{w} = \sum_{u' | u < u'} C_{u,w}^{w} \times x_{o^{u'}, j, u', v'}^{w-1} \]  
\forall w \in W, \forall n = 2, 3, \ldots, N, \forall u \in \{ (o^{w}, j, u, v) \in \Psi^{p,r^{w}} \}  
(1.5)

\[ C_{u,w}^{w} = \sum_{m=1}^{u} c_{d}^{w} q_{m}^{w} \left( u - \frac{m - 1 + m}{2} \right) \]  
\forall w \in W, \forall u \in \{ (o^{w}, j, u, v) \in \Psi^{p,r^{w}} \}  
(1.6)

With the calculation of passenger schedule delay costs, now we formally present the passenger side optimization problem. The objective of the problem, (2.1), is to minimize the sum of 1) total passenger schedule delay cost and 2) train layover cost at stations. For the first part in (2.1), we relate schedule delay costs of passengers boarding each subtrain to the starting arcs of the subtrain, and then sum over all subtrains. Note that this is a quadratic objective function as \( x \) variables are multiplied by \( C \) variables, which as shown above can themselves involve \( x \). For the second part, recall that each passenger subtrain \( r \) arriving at its destination is assigned one unit of time for sinking and \( t_{min}^{r} \) for minimum layover. The second term in (2.1) thus penalizes subtrain \( r \) for staying longer than \( (t_{min}^{r} + 1) \). Note that if layover at a station is fixed, i.e., \( t_{min}^{r} = t_{max}^{r} \), then the second component in (2.1) can be dropped.

\[ \min \sum_{w \in W, n=1,2,3,\ldots,N} \left( C_{u,w}^{w} + C_{u,w}^{w} \right) x_{o^{w}, j, u,v}^{w} + \sum_{(r,t) \in \mathbb{Z}^{p}, (t,f) \in \mathbb{Z}^{p}} c_{r}^{f} \left( f - t - (t_{min}^{r} + 1) \right) y_{r,t}^{r,f} \]  
(2.1)
The minimization problem is subject to a set of linear network and side constraints, as shown in (2.2)-(2.7) and (3.1)-(3.4):

**linear network constraints**

\[
\sum_{(o^c, j, u, v) \in \Psi_{P, N}^r} x^r_{o^c, j, u, v} = 1 \quad \forall n = 1, 2, \ldots, N, \forall w \in W
\]  \hspace{1cm} (2.2)

\[
\sum_{(d^n, e, u, v) \in \Psi_{P, N}^r} x^r_{d^n, e, r, u, v} = 1 \quad \forall n = 1, 2, \ldots, N, \forall w \in W
\]  \hspace{1cm} (2.3)

\[
\sum_{(a, i, u, t) \in \Psi_{P, r}} x^r_{a, i, u, t} = \sum_{(i, j, t) \in \Psi_{P, r}} x^r_{i, j, t, v} \quad \forall r \in R^p, i \in \{B | i \neq o^c\}, t \in T
\]  \hspace{1cm} (2.4)

\[
\sum_{(d^n, e, r, u, t) \in \Psi_{P, r}} x^r_{d^n, e, r, u, t} = \sum_{(t, \hat{r}) \in \Psi_{P, r}} y^r_{t, \hat{r}} \quad \forall r, \hat{r} \in R^p, (r, \hat{r}) \in Z^p, t \in T
\]  \hspace{1cm} (2.5)

\[
\sum_{(o^c, j, u, v) \in \Psi_{P, r}} x^r_{o^c, j, i, v} = \sum_{(t, \hat{r}) \in \Psi_{P, r}} y^r_{t, \hat{r}} \quad \forall r, \hat{r} \in R^p, (r, \hat{r}) \in Z^p, t \in T
\]  \hspace{1cm} (2.6)

\[
x^r_{i, j, u, v} y^r_{t, \hat{r}} \in \{0, 1\}
\]  \hspace{1cm} (2.7)

**side constraints**

\[
\sum_{r \in R^{P, N}} x^r_{i, j, u, v} \leq b^i_t \quad \forall i \in B, t \in T
\]  \hspace{1cm} (3.1)

\[
\sum_{r \in R^{P, S}} x^r_{i, j, u, v} + \sum_{v \in \{t+1, \ldots, t+1+\delta\}} x^r_{i, j, u, v} \leq k^i_t \quad \forall (a, t) \in T
\]  \hspace{1cm} (3.2)

\[
\sum_{r \in R^{P, N}} x^r_{a, j, u, v} + \sum_{v \in \{t+1, \ldots, t+1+\delta\}} x^r_{a, j, u, v} \leq b^i_t \quad \forall i \in B, t \in T
\]  \hspace{1cm} (3.3)

\[
\sum_{r \in R^{P, S}} x^r_{a, j, u, v} + \sum_{v \in \{t+1, \ldots, t+1+\delta\}} x^r_{a, j, u, v} \leq b^i_t \quad \forall i \in B, t \in T
\]  \hspace{1cm} (3.4)

In the linear network constraints, (2.2) and (2.3) guarantee unique departure and arrival for each subtrain. Note that it is sufficient to write constraints (2.2)- (2.3) just for the first subtrain of each physical train, as flow conservation constraints will guarantee unique departure of all remaining subtrains. However, establishing these constraints for all subtrains generates extra cuts, which help lower computation time. Flow conservation constraint (2.4) ensures existence of a subtrain path at each node and time. Constraints (2.5) and (2.6) establish the linkage between a pair of subtrains at the destination and origin stations. They together can also be viewed as flow conservation at the artificial sink nodes. Constraint (2.7) indicates that both x and y are 0-1 binary variables. Note that the unique departure of a subtrain from its origin (i.e., constraint (2.2)) and train path continuity (i.e., constraint (2.4)) guarantee the unique arrival of the subtrain at its destination. Hence, constraint (2.3) is not essential. However, we still keep the constraint because again it can generate extra cuts and help reduce computation time.

For the side constraints, (3.1) regulates capacity limit for each block. The capacity of a single block equals one, and the presence of a siding adds one more unit of capacity. \(b^i_t\) has a time index which enables time-dependent capacity variations (e.g., siding closure). Constraint (3.2) recognizes that the number of train transitions (in both directions) through a cell should not exceed the capacity of the cell at \(t\), for an interval defined by leading and lagging transition margins \(\delta\) and \(\epsilon\), \([t+1-\epsilon, \ldots, t+1+\delta]\). Constraints (3.3) and (3.4) manage headway, i.e., the minimum physical separation distance (measured in track blocks) between a pair of
consecutive subtrains, in two directions. Among these constraints, (2.2)-(3.1) are essential in train scheduling models and commonly used in the literature. Constraints (3.2)-(3.4) are borrowed from Harrod (2011).

In solving for the above program, the order of subtrains must be preserved, i.e., later departure of subtrain \(r_{n+1}^w\) than of subtrain \(r_n^w, n = 2, ..., N\) is not allowed. To ensure this, we add to (2.1) a penalty term (4) which represents combinations of neighboring subtrains that violate the train order. Each combination is multiplied by a big number \(M\). In this study we consider \(M\) to be the sum of maximum possible schedule delay of each traveler, which equals the time difference between one’s PDT and the beginning or end of the planning horizon, whichever gives a greater schedule delay. Note that if we consider passenger trains to have different speed but do not allow overtaking (as is often the case in the US), then a similar big \(M\) term penalizing violation of subtrain orders at the arrival end should also be added. Note that when (4) is added to (2.1), it automatically takes care of the \(EADT\) and \(LAAT\) for each train. Therefore we do not need to specify \(EADT\) and \(LAAT\) separately.

\[
\sum_{n=2}^{N} M \times X_{o_{w,f,i,u,v}}^{r_{n-1}} \times X_{o_{w,i,u,v}}^{r_{n}}
\]

(4)

3.2.2 Lower level: Freight train scheduling

The lower level problem consists in developing freight train schedules, conditional by passenger train schedules generated from the upper level. Here we do not need to decompose each physical freight train into subtrains, as freight trains do not stop at intermediate stations. However, en-routing stopping is possible when encountering capacity constraints.

In the US, a freight train is dispatched whenever the train receives enough load (Cordeau et al., 1998). A freight train generally prefers to depart as early as possible, which increases the probability of on-time arrival and makes more capacity available to subsequent freight trains. The freight operator also wants to minimize en-route train delays which incur additional operating cost. Furthermore, a loss of demand and consequent cost would occur when freight demand is more than the freight operator can handle given the available line capacity. For strategic planning purposes, in this study we consider constant train speed and rolling stock management cost. Therefore, the cost minimization problem consists in minimizing the sum of costs due to train departure delay, en-route delay, and lost demand:

\[
\text{Min} \sum_{r \in R} \left( c_e^T(u - EADT^r) - c_p^T \right) X_{o_{w,i,u,v}}^{r} + \sum_{r \in R} c_p^T X_{o_{w,i,u,v}}^{r}
\]

(5.1)

s.t.

\[
\sum_{r \in R} X_{o_{w,i,u,v}}^{r} \leq 1 \quad \forall \{r \in R'\}
\]

(5.2)

\[
\sum_{r \in R} X_{d_{w,e,u,v}}^{r} \leq 1 \quad \forall \{e \in E\}
\]

(5.3)

\[
\sum_{r \in R} X_{d_{i,u,t}}^{r} = \sum_{r \in R} X_{i,j,t,v}^{r} \quad \forall \{r \in R', i \in \{B| i \neq o_{w}^t\}, t \in T\}
\]

(5.4)

\[
\sum_{r \in R} X_{d_{w,e,u,v}}^{r} = \sum_{r \in R} X_{o_{w,i,u,v}}^{r}
\]

(5.5)

\[
\{2.2\}-\{2.7\} \quad \forall \{r \in R', \text{given } EADT^r \text{ and } LAAT^r \text{ found from the upper level}\}
\]

(5.6)

\[
\{3.1\}^{'}-\{3.4\}^{'} \quad \forall \{r \in R, \text{given } EADT^r \text{ and } LAAT^r \text{ found from the upper level}\}
\]

(5.7)

\[
x_{i,j,u,v}^{r} \in [0,1]
\]

(5.8)

The first term in (5.1), \(c_e^T(u - EADT^r)\), denotes freight train cost due to departure delay. Given a prescribed maximum number of freight trains that desire to be scheduled (i.e., total freight train demand), cost due to lost demand is represented by the negative of the number of freight trains that can be actually scheduled, times the
cost associated with one train of lost demand, \( c^p_r \). This means that, the more freight trains that can be accommodated, the smaller the lost demand cost. The last term in (5.1) captures en-route delay costs, which are associated with stopping hyperarcs \((i = j)\).

The cost minimization is subject to linear network and side constraints. Constraints (5.2) and (5.3) enforce unique departure and arrival for each freight train. In contrast to (2.2) and (2.3), we consider inequalities as a freight train may not be scheduled due to capacity constraints. Constraint (5.4) guarantees the continuity of each freight train path. Constraint (5.5) is optional which equates the number of non-zero starting arcs in one direction with the number of non-zero sinking arcs in the opposite direction. Having (5.5) thus results in equal numbers of freight trains running in the two directions. Constraint (5.6) imposes passenger train schedules, by setting \( EADT^r \) and \( LAAT^r \) for passenger subtrains equal to what is found from the upper level problem. In this way the feasible path arcs of passenger subtrains will be identical to the optimal path arcs from solving the upper level problem. Constraint (5.7) governs block occupancy, transition, and headway constraints for all trains, in a similar way as in (3.1)-(3.4), except that now we consider all trains (i.e., \( \forall r \in R \)) and passenger train schedules are given. Note that an alternative approach to using \( EADT^r \) and \( LAAT^r \) is to reduce block and cell capacity \( b^l_1 \) and \( k^o_1 \) by considering the occupancy of passenger subtrains from the upper level. Constraint (5.8) regulates that the \( x \) variables are binary.

### 4 Solution strategies

Our investigation of solution strategies focuses on reformulating the upper level problem to improve computational performance. For the lower level, the linear integer program can be conveniently solved using off-the-shelf techniques. As mentioned in sub-Section 3.2.1, the objective (2.1) at the upper level is a quadratic function of \( x \), making the problem a Quadratic Integer Program (QIP). An abstract expression of the QIP is shown as (6.1)-(6.4). QIPs are in general NP-hard, and remain so after relaxing the integer requirement for \( x \), if matrix \( H \) in (6.1) is indefinite (Pardalos and Vavasis, 1991; Sahni, 1974). Our numeric analyses reveal that \( H \) always have both positive and negative eigenvalues. Hence, the upper level problem falls into the category of NP-hard programs.

\[
\begin{align*}
\text{Max } f \cdot x & + \frac{1}{2} x^T H x \\
\text{s.t.} & \quad A_{\text{ineq}} x \leq b_{\text{ineq}} \\
& \quad A_{\text{eq}} x = b_{\text{eq}} \\
& \quad x \in \{0,1\}
\end{align*}
\]  
(6.1)

We consider three possible remedies that either partially or full linearize the original problem. The first remedy drops the big \( M \) term (4) from the objective function. Instead, a new set of constraints (7) is added to ensure subtrain orders (i.e., the departure of \( r^{w-1}_n \) is no later than the departure of \( r^n_w \)). The reason for dropping the big \( M \) is that it creates large difference in value – or even in magnitude – between this and other terms in the objective function. This can lead to “round-off error in floating point calculations which makes it difficult for the algorithm to distinguish between the error and a legitimate value” (Klotz and Newman, 2013). In contrast to this, adding constraints (7) in place of the big \( M \) will introduce new cuts which help improve computational efficiency.

\[
\begin{align*}
\sum_{(o,w,j,u,v) \in \Psi^p_{r^{w-1}_n}} u' \ast x^{r^{w-1}_n}_{o,w,j,u,v}' & \leq \sum_{(o,w,j,u,v) \in \Psi^p_{r^n_w}} u \ast x^{r^n_w}_{o,w,j,u,v} \\
\forall w & \in W, \{ \forall n = 2, 3, ..., N \}
\end{align*}
\]  
(7)

A second remedy is to completely linearize the QIP by replacing each feasible combination of starting arcs of two consecutive subtrains in (2.1) by a binary \( z \) variable, as shown in (8.1), and by further adding constraints (8.2)-(8.4). Constraints (8.2) and (8.3) stipulate that if either \( x \) variable in \( z \) is zero then \( z \) must be zero. Constraint (8.4) ensures that \( z \) equals one if and only if both \( x \)'s equal one.
\begin{align}
& z_{u',u}^{w,n-1} = x_{o',j,u', v'}^{w,n-1} x_{o',j,u,v}^{w,n} \\
& \forall w \in W, \{\forall n = 2, 3, \ldots, N\}, \{\forall (u', u) | u' < u, (o', j, u', v') \in \Psi^{p,n-1}, (o', j, u, v) \in \Psi^{p,n}\} 
\end{align}

(8.1)

\begin{align}
& z_{u',u}^{w,n-1} \leq x_{o',j,u', v'}^{w,n-1} \\
& \forall w \in W, \{\forall n = 2, 3, \ldots, N\} 
\end{align}

(8.2)

\begin{align}
& z_{u',u}^{w,n-1} \leq x_{o',j,u,v}^{w,n} \\
& \forall w \in W, \{\forall n = 2, 3, \ldots, N\} 
\end{align}

(8.3)

\begin{align}
& x_{o',j,u', v'}^{w,n-1} + x_{o',j,u,v}^{w,n} \leq 1 + z_{u',u}^{w,n-1} \\
& \forall w \in W, \{\forall n = 2, 3, \ldots, N\} 
\end{align}

(8.4)

Our computational experiments show that the above linearization still cannot solve large problems within a reasonable amount of time. This is because linearization introduces many new binary variables and constraints, thus substantially increasing the problem size. On the other hand, we note that passenger subtrains are all "must-run". As a result, one and only one of the starting arcs for each subtrain is equal to one. Translating this to \(z\), it means that one and only one of the \(z\)'s associated with each pair of consecutive subtrains is equal to one. This leads to our third remedy, which retains constraints (8.2)-(8.3) but replaces (8.4) by (9). Constraint (9) has fewer constraints with equalities, which make the branching strategy easier. In fact, each time when one of the \(z\)'s in (9) is fixed to one, all other \(z\)'s in the same equation automatically become zero. This improves the search tree.

\begin{equation}
\sum_{(u',u) | u' < u, (o',j,u',v') \in \Psi^{p,n-1}, (o',j,u,v) \in \Psi^{p,n}} z_{u',u}^{w,n-1} = 1
\end{equation}

\(\forall w \in W, \{\forall n = 2, 3, \ldots, N\}\)

Summing up, four model formulations are possible for the upper level problem:

\textbf{F1:} \quad \textit{Min} \ (2.1) + (4) \\
\textit{S. t.} \ (2.2)-(2.7) \\
\quad (3.1)-(3.4) \quad (10)

\textbf{F2:} \quad \textit{Min} \ (2.1) \\
\textit{S. t.} \ (2.2)-(2.7) \\
\quad (3.1)-(3.4) \\
\quad (7) \quad (11)

\textbf{F3:} \quad \textit{Min} \ (2.1)' \\
\textit{S. t.} \ (2.2)-(2.7) \\
\quad (3.1)-(3.4) \\
\quad (7) \\
\quad (8.2)-(8.4) \\
\quad z \in \{0,1\} \quad (12)

\textbf{F4:} \quad \textit{Min} \ (2.1)' \\
\textit{S. t.} \ (2.2)-(2.7) \\
\quad (3.1)-(3.4) \\
\quad (7) \\
\quad (8.2)-(8.3) \\
\quad (9) \\
\quad z \in \{0,1\} \quad (13)

where (2.1)' denotes a reformulation of (2.1) by substituting the product of \(x\)'s by \(z\)'s, as defined by (8.1).
5 Numerical analysis

This section demonstrates the implementation of the two-level train scheduling models. We start with determining values for model parameters. We then solve a sample problem, which includes investigating the impact of train speed heterogeneity and freight side delay tolerance levels. A simplified case study of the Chicago-St Louis higher speed rail corridor is presented in the end. Mathematical programs are coded in MATLAB (2013b) and solved using IBM ILOG CPLEX Optimizer V12.6 on a computer with Intel Core i7 3770 3.4 GHz CPU and 12GB of RAM.

5.1 Model parameters

Implementing the model requires a number of parameters. On the passenger side, these include the passenger PDT profiles and the unit value for passenger schedule delay ($c_d^p$ and $c_d^s$). Currently, the literature on passenger PDT is still limited. In this study we use an existing passenger PDT profile from Cascetta and Coppola (2012) (Figure 6), and further assume the shape of passenger PDTs to be identical for all ODs. For value of passenger schedule delay, we follow Corman and D’Ariano (2012), who specifically estimate values of passenger travel time for rail, and consider passenger value of schedule delay to be 1.5 times of line-haul travel time. We use the reference values for line-haul travel time provided by the US Department of Transportation (DOT) for rail, which equates $32.9/\text{hr}$ and $58.9/\text{hr}$ for personal and business trips (US DOT, 2011). We consider 7% of total trips for business purposes, following Illinois DOT (2012). Consequently, the passenger value of schedule delay (assuming $c_d^p = c_d^s$) is $52/\text{hr}$.

The freight side involves three unit cost parameters related to: departure delay ($c_e^f$), en-route delay ($c_r^f$), and lost demand ($c_l^f$). As train operating cost due to departure delay is not likely to be significant (delayed trains would simply wait at the departure point), we consider $c_e^f$ to be equal to freight value of time (VOT). We use the freight VOT estimate from Vieira (1992) (Feo-Valero et al., 2011; Zamparini and Reggiani, 2007) and convert it to 2010 value using a 3% inflate rate. The resulting freight VOT is $0.955/\text{ton-hr}$. We multiply this value by the average tonnage hauled by a typical freight train in the US (3585 tons/train, based on AAR (2012)). The estimated $c_e^f$ is equal to $3424.1/\text{train-hr}$.

The unit cost due to en-route delay $c_r^f$ is expected to be greater than $c_e^f$, because trains will be in operation (although little fuel consumption will incur while stopping en route). We consider $c_r^f$ to be the sum of freight VOT and average train operating cost excluding fuel cost. The average train operating cost is obtained by dividing total operating expenses by the total train-hours over all Class I railroads in the US (AAR, 2012). We discount the operating cost by 10% to reflect the exclusion of fuel cost (Mizutani, 2004). This leads to $163.1/\text{train-hr}$. Further adding freight VOT, $c_r^f$ is estimated to be $5057.2/\text{train-hr}$.

Explicit estimates of $c_l^f$ do not exist in the US. Johnson and Nash (2005) point out that estimating the cost to lost rail freight demand requires revealing the value placed on the slot by an alternative user. The estimated opportunity cost of a slot for freight trains is £0.0149/ton-km in the UK, or $0.0566/\text{ton-mile}$ in 2010 value.
(using 3% inflation rate and a pound-dollar exchange rate of 1.66) (MeasuringWorth, 2013). We follow the same idea and approximate lost demand cost using average railroad revenue. In 2010 US Class I Railroads hauled 1,691 billion ton-miles of freight with a total revenue of $58.41 billion (AAR, 2012), resulting in an average operating revenue of $0.035/ton-mile. This is very close to Johnson and Nash’s estimate. Multiplying $0.035/ton-mile by the average tonnage hauled by each train, the estimated cost due to lost demand is $123.8/mile. \( c_p^* \) is obtained by multiplying $123.8/mile by the length of the corresponding train trip.

### 5.2 A Sample problem

The sample problem considers a single-track line with 11 blocks. There is no intermediate station along the line. Therefore only two OD pairs are involved, one for each direction. Each physical passenger train is equivalent to a single subtrain. Six out of the 11 blocks are identical single-track segments. The other five blocks, inserted evenly between the single-track segments, are shorter, identical double-track segments, i.e., single-track segments with a siding. Each single-track segment is 18 miles long (except the first and last segments which are 19 miles long). Each double-track segment siding is 2 miles long. This problem size is comparable to the ones considered in Harrod (2011).

We consider train scheduling between 5:00 AM and 9:30 PM. Passenger trains run at 120 mph with no stops en-route. In Appendix 2, we show that allowing passenger trains to stop en-route insignificantly influences the results of the small problem, as well as those of the larger problem in subsection 5.5. When six passenger trains run each way, we assume a daily demand of 760 and 710 passengers in the two directions. Total passenger demand is elastic with respect to the number of trains. We follow Adler et al. (2010) and use a frequency elasticity of 0.4 to estimate total passenger demand with lower train frequencies. While there can be alternative approaches to deal with demand variations in response to train schedule changes (e.g., Talebian et al. (2015)), in this study we proportionately adjust the passenger PDT profile when total demand changes.

We assume that freight trains run at 60 mph. The EADT of freight trains are assumed uniformly distributed with one-hour intervals. The EADT of the first freight train is at the beginning of the planning horizon. The maximum delay tolerance for a freight train is 50% of the train’s non-stop travel time (Sahin et al., 2008). LAAT for each freight train is obtained using EADT and delay tolerance information. Same as for passenger subtrains, LAAT and EADT are used to generate the set of feasible freight train path arcs. With no other constraints, a maximum of 15 freight trains can be scheduled in each direction (noting that it takes two hours for a freight train to run end-to-end on the line). The minimum headway \( h^r (\forall r \in R) \) is one block following Harrod (2011).

The choice of time resolution can significantly affect the computation time. According to Sahin et al. (2008), considering 5-min intervals provides sufficient accuracy for strategic level train scheduling. We follow this suggestion and choose five minutes as the length of a time unit. On the other hand, passenger PDT profiles are usually identified at coarser intervals, in our case each with 15 minutes. To reconcile the different time resolutions, we assume that passenger PDTs are uniformly distributed in each 15-minute interval. Thus the numbers of passengers in each 5-minute interval is one third the numbers of passengers in the corresponding 15-min period.

We consider all four formulations F1-F4 while solving the passenger side problem. F1 shows very limited capability, failing to solve any passenger train scheduling problem due to memory overflow. In the scenario with two passenger trains in each direction, F2 yields the optimal solution in 46.25 hours. Computation is significantly improved after the problem is fully linearized (F3 and F4). In particular, the optimal solution using F4 can be found within a minute for up to six passenger trains running in each direction (Table 2). In what follows, we always consider F4 to solve the passenger side problem. On the freight side, optimal solutions can be found under 10 seconds in all cases investigated.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Passenger train in each direction</th>
<th>Solving time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sc. 1</td>
<td>2 passenger trains</td>
<td>6 sec</td>
</tr>
<tr>
<td>Sc. 2</td>
<td>3 passenger trains</td>
<td>16 sec</td>
</tr>
<tr>
<td>Sc. 3</td>
<td>4 passenger trains</td>
<td>25 sec</td>
</tr>
<tr>
<td>Sc. 4</td>
<td>5 passenger trains</td>
<td>33 sec</td>
</tr>
<tr>
<td>Sc. 5</td>
<td>6 passenger trains</td>
<td>44 sec</td>
</tr>
</tbody>
</table>
Table 3 reports the departure time of passenger trains and passenger demand with different numbers of trains. When the number of trains is greater than one, the morning and afternoon peaks will each have one train scheduled. Detailed train schedules in time-space diagrams are provided in Appendix 3. Given the same shape of passenger PDTs for the two ODs, symmetric train departures are expected. Therefore, two opposing trains departing at the same time will meet at the midpoint, where a siding exists. The only asymmetry occurs when one train is scheduled each way, which is due to rounding passenger numbers.

<table>
<thead>
<tr>
<th>Passenger trains in each direction</th>
<th>Passengers</th>
<th>Passenger train departure time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NB</td>
<td>SB</td>
</tr>
<tr>
<td>1</td>
<td>381</td>
<td>407   (14:25,12:15)</td>
</tr>
<tr>
<td>2</td>
<td>476</td>
<td>509   (8:10,8:10)</td>
</tr>
<tr>
<td>3</td>
<td>549</td>
<td>587   (8:10,8:10)</td>
</tr>
<tr>
<td>4</td>
<td>610</td>
<td>652   (7:25,7:25,9:35)</td>
</tr>
<tr>
<td>5</td>
<td>663</td>
<td>709   (7:20,7:20,9:20)</td>
</tr>
<tr>
<td>6</td>
<td>710</td>
<td>760   (7:15,7:15,9:10)</td>
</tr>
</tbody>
</table>

Table 3: Train departure times and passenger demand

Figure 7 plots total and per passenger schedule delay cost as a function of the number of trains. Both follow a decreasing trend as more trains are scheduled. However, the extent of cost reduction diminishes as more trains are scheduled. When six passenger trains are scheduled, average passenger schedule delay cost is $30, which is comparable to in-vehicle travel time cost ($34.7/hr * 1 hr = $34.7).

![Figure 7: Total and per passenger schedule delay cost](image)

Table 4 documents passenger schedule delay cost for each train and the number of passengers on each train. Because passenger PDT fluctuates in a day, different trains are expected to have uneven numbers of passengers onboard and incur non-uniform schedule delay costs. We observe that trains with departures closer to morning and evening peaks carry more passengers, consequently having greater share of passenger schedule delay despite smaller schedule delay per passenger during those peaks. For instance, in the six-train scenario, schedule delay costs associated with train 1/2 and train 3/4, which depart during the morning peak, are about twice as those for train 5/6. Recall that we consider trains to have sufficient seat capacity for strategic level planning. Even if we assume 378 seats on a train, following the suggested 5-car train configuration by Illinois Department of Transportation (IDOT, 2012), passengers can be fully accommodated under all scenarios except when only one train is scheduled in each direction.4

---

4 If a large number of passengers were unattended (i.e., passenger train seating capacity constraints are severe), then an alternative modeling for the upper-level problem could be to minimize the total cost of passenger schedule delay and unattended demand.
Table 4: Passenger schedule delay cost (in $000) and passengers onboard (in parenthesis) for each train in various scenarios

<table>
<thead>
<tr>
<th>Number of daily passenger trains in each direction</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trains in direction 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 1</td>
<td>85.75 (381)</td>
<td>11.16 (230)</td>
<td>13.49 (258)</td>
<td>5.10 (158)</td>
<td>4.65 (150)</td>
<td>4.75 (157)</td>
</tr>
<tr>
<td>Train 3</td>
<td>14.27 (246)</td>
<td>5.79 (146)</td>
<td>5.66 (134)</td>
<td>4.17 (135)</td>
<td>4.46 (150)</td>
<td></td>
</tr>
<tr>
<td>Train 5</td>
<td>4.92 (145)</td>
<td>6.95 (160)</td>
<td>3.19 (74)</td>
<td>2.35 (63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 7</td>
<td>5.90 (158)</td>
<td>5.29 (154)</td>
<td>2.83 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 9</td>
<td></td>
<td>5.83 (150)</td>
<td>2.93 (124)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 11</td>
<td></td>
<td></td>
<td>3.87 (116)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trains in direction 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 2</td>
<td>93.56 (407)</td>
<td>14.33 (255)</td>
<td>14.58 (284)</td>
<td>6.03 (180)</td>
<td>5.61 (174)</td>
<td>5.40 (173)</td>
</tr>
<tr>
<td>Train 4</td>
<td>16.34 (254)</td>
<td>5.59 (144)</td>
<td>5.98 (141)</td>
<td>4.18 (137)</td>
<td>4.71 (158)</td>
<td></td>
</tr>
<tr>
<td>Train 6</td>
<td></td>
<td>6.85 (161)</td>
<td>3.37 (79)</td>
<td>2.45 (67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 8</td>
<td></td>
<td>6.63 (170)</td>
<td>5.48 (161)</td>
<td>2.93 (105)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 10</td>
<td></td>
<td></td>
<td>6.31 (158)</td>
<td>3.15 (132)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 12</td>
<td></td>
<td></td>
<td></td>
<td>4.24 (125)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incremental passenger benefits ($000)</td>
<td>151.23</td>
<td>13.69</td>
<td>6.17</td>
<td>5.09</td>
<td>7.2</td>
<td></td>
</tr>
</tbody>
</table>

The last line in Table 4 provides estimated incremental passenger benefits. Under the assumption of a linear demand function or small changes in passenger generalized cost, the “rule of one-half” can be employed to estimate incremental passenger benefits (Small and Verhoef, 2007). The passenger generalized cost is the sum of rail fare and schedule delay cost. Since rail fare is assumed to be constant, schedule delay cost is the only variable component. Figure 8 plots the relationship between the average schedule delay cost per passenger and total passenger demand. The relationship is very close to linear, except for the case with one passenger train. Thus following the rule of one-half the incremental passenger benefits from adding a train can be approximated by $\frac{1}{2} \Delta_{SD} \ast (D_0 + D_1)$, where $\Delta_{SD}, D_0,$ and $D_1$ denote the change in average schedule delay cost per passenger and passenger demand before and after adding a train. This is illustrated in Figure 8, where the hatched area represents the incremental passenger benefits of increasing the number of trains from two to three. The rectangular area represents the benefits to existing passengers due to increasing the number of trains (and consequently reduced average schedule delay), and the triangular area corresponds to the benefits to new (induced) travelers. Adding these two parts together, total incremental passenger benefits equals the trapezoid area. We apply this to each station pair and sum over the two station pairs to obtain the total incremental benefits.
Figure 8: Relationship between average passenger schedule delay cost and total passenger demand and an illustration of incremental passenger benefits of increasing the number of passenger trains

Figure 9 presents the freight side costs against the number of passenger trains. Recall that freight side cost consists of three parts: costs due to lost demand, late departure, and en-route delay. As expected, total freight side cost increases with the number of passenger trains. This is particularly true for costs due to lost demand and en-route delay. When the number of passenger trains is less than three, all freight trains can run on the line. If 3-5 passenger trains are scheduled each way, two freight trains will be forced to be out of service. The lost demand increases to four freight trains when six passenger trains are scheduled. In the latter case, cost due to lost demand becomes the most important component in the total.

We further compare marginal cost reduction on the passenger side with marginal cost increase on the freight side when increasing the number of passenger trains. Figure 10 shows that net marginal benefit gain only occurs when passenger trains increase from one to two in each direction.

Figure 9: Freight-side cost as a function of the number of passenger trains ($000)
5.3 The impacts of speed heterogeneity

Speed heterogeneity brought by mixed train operations is an important issue on shared-use rail corridors (Dingler et al., 2009; Dingler, 2010; Sogin et al., 2013). In this subsection we fix passenger train speed but vary the speed of freight trains by a wide range from 12 mph to 120 mph. Figure 11 presents the sum of freight side cost under different speeds, as a function of the number of passenger trains. Each line corresponds to one freight train speed. We observe two general trends. First, increasing freight train speed, which reduces the speed difference between passenger and freight trains, leads to smaller freight side cost. This is consistent with findings in the literature (e.g. Dingler et al., 2009). Two exceptions are freight trains running at 120 mph and 75 mph with five and six passenger trains. A closer look into train schedules reveals that the allowed operating time intervals for some freight trains have greater overlap with passenger train paths at 120 mph than at 75 mph, leading to infeasible operation of those freight trains. Second, adding passenger trains increases freight side cost, with only two exceptions: one on the 12-mph curve with one passenger train; the other on the 120-mph curve with five passenger trains. The anomalies might be attributed to better forms of available capacity to accommodate freight trains after altering passenger train departures.

The above conjecture is further confirmed by lost demand under different freight train speeds (Figure 12). When there are five or six passenger trains, some freight trains cannot be scheduled at 120-mph speed. In contrast, all freight trains can be accommodated at 75-mph speed. By and large, the number of freight trains shows a non-increasing trend with the number of passenger trains. If freight trains run at 12 mph, freight service will disappear entirely if scheduling four or more passenger trains.

---

5 It is likely that train operating cost while running will be different when speed changes. However, our analysis focuses on the three infrastructure capacity related cost components.
Figure 11: Total freight side cost with different freight train speeds

Figure 12: The number of freight trains scheduled with different freight train speeds

Figure 13 and 14 present en-route and departure delay costs of freight trains, in total and average per running train, as a function of freight train speed and the number of passenger trains. When freight trains run at 120 mph, which is the same speed as passenger trains, we observe no en-route delay. Average en-route delay cost per running train exhibits a consistent pattern – it increases when lowering freight train speed – with two exceptions with zero and two passenger trains at 12-mph speed. No other general finding can be drawn for en-route delay, however. Train departure delay cost has an even less clear trend in both total and average.
5.4 The impact of freight train delay tolerance levels

This subsection examines the sensitivity of freight side cost to train delay tolerance. Previously we use 0.5 times the unimpeded trip time to bound the sum of departure and en-route delays for freight trains. In this subsection we vary this multiplier from 0.1 and 0.6 with 0.1 increments. Our numeric experiments show that no further improvement in freight side cost is found if the multiplier is greater than 0.6.

Figure 15 illustrates lost freight demand and associated cost under different levels of delay tolerance. As expected, greater delay tolerance leads to non-decreasing numbers of running freight trains. Consistent with Figure 12, increasing the number of passenger trains generally reduces the number of running freight trains and increases lost demand cost across all delay tolerance levels. For example, if only one passenger train runs each way, no lost demand will occur with delay tolerance greater than 0.2. When six passenger trains are scheduled, a delay tolerance of 0.5 or 0.6 will incur $50000 lost demand costs, which is about one sixth of the cost with 0.1 delay tolerance.

The savings in lost demand due to greater delay tolerance should be weighed against the degradation of train on-time performance. Figure 16 shows that both departure and en-route delay costs become larger as we
progressively relax delay tolerance: when a 0.1 multiplier is used there is no train departure delay. En-route delay also remains minimum. Maximum departure and en-route delay costs are achieved with 0.6 tolerance level. Note that in most cases delay related costs are much smaller than the cost due to lost demand, which suggests that tolerating more delays could reduce overall freight side cost. This is in line with the fact that freight railroads are typically more willing to bear high delays than to cut operations. The sensitivity of delay costs to delay tolerance varies by the number of passenger trains. The difference between maximum and minimum en-route delay costs generally increases with the number of passenger trains scheduled. In contrast, the variation of departure delay costs shows a less clear trend when adding more passenger trains.

![Graph showing the number of freight trains scheduled and corresponding lost demand cost for various levels of freight train delay tolerance.]

**Figure 15:** The number of freight trains scheduled and corresponding lost demand cost for various levels of freight train delay tolerance

![Graph showing freight en-route and departure delay costs under different delay tolerance levels.]

**Figure 16:** Freight en-route and departure delay costs under different delay tolerance levels

### 5.5 A larger problem

In this subsection we consider applying the model for strategic level train planning on a simplified Chicago-St Louis shared-used rail corridor. The track of the corridor is owned by Union Pacific railroad and currently being upgraded to accommodate 110-mph passenger rail service. The line schematic is shown in Figure 17
which is based on Harnish (2013). The corridor is 285 miles long and has 14 single-track and 17 double-track segments. This is comparable to problem sizes in Li et al. (2008), Burdett and Kozan (2009), Dündar and Şahin (2013), and Li et al. (2014). Although the prospective passenger service stops at nine stations (including Chicago and St Louis), 90% of OD trips will be among six stations, i.e., Chicago, Joliet, Normal, Springfield, Alton, and St. Louis (IDOT, 2012). In our analysis only these six stations are considered. We further assume that passenger train layover time at intermediate stations is equal to one time unit (in our model this equals 5 min).

![Figure 17: Schematic of the Chicago – St. Louis shared-used rail corridor](image)

We consider near-term (2015) and longer-term (2020) scenarios using passenger demand forecast from IDOT (2012). For the 2015 scenario, Table 5 shows passenger OD demand based on aggregate demand forecast and the existing demand distribution among ODs. At each station, we sum all originating passengers traveling in one direction and distribute them over the course of a day using Figure 6 to construct passenger PDT. IDOT (2012) suggests five daily services in each direction in the near term. Given the six stations along the line, each physical train is decomposed into five subtrains. We consider an average speed of 90 mph to account for train acceleration and deceleration.

On the freight side, we assume that freight trains run at 30 mph. In current traffic, very few freight trains traverse the entire corridor, with more freight trains near the south end (IDOT, 2012). In view of the actual traffic, we consider the following freight demand: 8 trains between Chicago and Joliet; 6 train between Joliet and Normal; 4 train between Normal and Springfield; 6 trains between Springfield and Alton; and 10 trains between Alton and St. Louis. The freight demand is assumed to be equal in both directions. Same as in sub-Section 5.2, the EADTs of freight trains for each OD are uniformly distributed over the planning horizon; the LAAT for each freight train equals its EADT plus 1.5 times of the unimpeded travel time.

<table>
<thead>
<tr>
<th>O\D</th>
<th>Chicago</th>
<th>Joliet</th>
<th>Normal</th>
<th>Springfield</th>
<th>Alton</th>
<th>St. Louis</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicago</td>
<td>0</td>
<td>7</td>
<td>286</td>
<td>215</td>
<td>58</td>
<td>271</td>
<td>837</td>
</tr>
<tr>
<td>Joliet</td>
<td>6</td>
<td>0</td>
<td>31</td>
<td>19</td>
<td>7</td>
<td>22</td>
<td>85</td>
</tr>
<tr>
<td>Normal</td>
<td>286</td>
<td>32</td>
<td>0</td>
<td>23</td>
<td>8</td>
<td>26</td>
<td>375</td>
</tr>
<tr>
<td>Springfield</td>
<td>212</td>
<td>19</td>
<td>23</td>
<td>0</td>
<td>21</td>
<td>43</td>
<td>318</td>
</tr>
<tr>
<td>Alton</td>
<td>57</td>
<td>7</td>
<td>9</td>
<td>23</td>
<td>0</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>St. Louis</td>
<td>268</td>
<td>23</td>
<td>27</td>
<td>45</td>
<td>4</td>
<td>0</td>
<td>367</td>
</tr>
<tr>
<td>Total</td>
<td>829</td>
<td>88</td>
<td>376</td>
<td>325</td>
<td>98</td>
<td>366</td>
<td>2,082</td>
</tr>
</tbody>
</table>

The upper level passenger train scheduling problem, which includes 421,627 variables, 52,881 equality constraints, and 755,138 inequality constraints, is solved in 15 minutes. On average, each passenger incurs 53.7 minutes of schedule delay, or $46.6 schedule delay cost. Compared to the proposed rail fare ($39) between Chicago and St. Louis (IDOT, 2012), the schedule delay cost is substantial. All passengers will be accommodated if we again consider a seating capacity of 378 passengers for each train. On the freight side, only one train between St. Louis and Alton and one train between Springfield and Normal cannot be accommodated. All train paths are shown in Figure 18. The breakdown of freight side cost is presented in the first bar of Figure 19.

---

6 The Texas Eagle service, which runs one train per day between Chicago and San Antonio, Texas through St Louis, and continues to Los Angeles (thus much longer than the Chicago – St Louis corridor), may still be using the Chicago – St Louis corridor after the 110-mph train service is in place. Further analysis could look into scenarios with the presence of this single-train service, for example, by assigning the train a predetermined schedule.
In the longer-term (2020) scenario, passenger OD trips are interpolated using IDOT 2015 and 2030 ridership forecast (IDOT, 2012), assuming a constant demand growth rate. Total passenger trips across all OD pairs presented in Table 5 will increase to 2272 per day. With the same number of passenger trains, average schedule delay cost remains $46.6/passenger, which is not surprising because demand is proportionately scaled. If one more passenger train is added in each direction, average schedule delay cost will be $44.3/passenger, or a $2.3 reduction.

The freight side traffic is also expected to increase over time. Based on IDOT’s projection (IDOT, 2012), we consider the following freight traffic for 2020: 16 trains between Chicago and Joliet; 20 trains between Joliet and Normal; 18 trains between Normal and Springfield; 22 trains between Springfield and Alton; and 24 trains between Alton and St. Louis. The resulting freight side costs and their breakdown are shown in the two right bars of Figure 19. The freight side will experience significant increase in all three cost components. In particular,
adding one more passenger train would lead to more than doubled cost to lost demand. This evidences severe capacity constraints given future passenger and freight demand.

## 6 Concluding remarks

This paper contributes to the literature by developing a strategic level train planning model for mixed operations on shared-use rail corridors. Given the access priority of passenger trains as mandated by the US federal law, we propose a two-level modeling framework to sequentially determine passenger and freight train schedules. Our model considers schedule delay for passengers and lost demand for the freight host railroad, both of which are often ignored but prove important. We adopt a hypergraph-based approach to explicitly account for potential path conflicts during train transitions between track blocks. In order to efficiently solve the upper level problem, we linearize the quadratic integer program formulation and further take advantage of the problem-specific model structure, which leads to a substantial improvement in computational performance.

Using realistic parameter values, our numeric experiments show that passenger cost due to schedule delay is comparable to in-vehicle travel time cost for short-distance travel. Scheduling more passenger trains on a shared-use corridor lowers passenger schedule delay, but the tradeoff is an increase in freight cost. The marginal freight cost increase is in most cases greater than the marginal reduction in passenger schedule delay cost. Of course, passenger schedule delay cost and its variation to the number of trains depend on the level and distribution of passenger demand; similarly, the changes in freight side cost hinge upon the current level of freight demand as well as passenger traffic. The heterogeneity of train speed reduces the number of freight trains that can run on a corridor. An increment of the allowed tolerance of freight train delays also reduces lost demand and the overall cost on the freight side. When applying the model to the Chicago – St Louis corridor, we find that severe capacity constraints would arise given demand increase in the near future. The modeling framework developed in this paper could be further applied to other scenarios with different parameter values.

The present research can be extended in several directions. First, although “unattended” demand is not significant in our numerical experiments, it can be important when passenger demand is high with respect to train seating capacity. In this case, how travelers’ train selection behavior will be affected and consequently how passenger benefits are assessed need to be investigated. Second, from the passenger train operator’s perspective, minimizing train operating cost and improving quality of service (i.e., minimizing passenger schedule delay) could be two separate objectives. It would be interesting to consider a multi-objective approach to determine passenger train schedules. Third, our model involves some simplifications which can be replaced with more detailed specifications. For instance, train acceleration and braking and speed variation in general are not explicitly considered in this paper. In addition, for other shared-use corridors there may exist multiple types of passenger services with different stops and overtaking possibilities. Extension of the current model could be directed to explicitly incorporating those aspects. Fourth, modified formulations of the model to encompass en-route delay of passenger trains and time that passengers spend at the origin and destination could offer alternative ways to characterize train schedule inconvenience to travelers.

### Acknowledgment

This research was partially supported by Illinois Department of Transportation’s Metropolitan Transportation Support Initiative (METSI) through the Urban Transportation Center at the University of Illinois at Chicago. We thank Professor Steven Harrod for sharing his experiences in developing hypergraph-based models. We are also very grateful to three anonymous reviewers and especially Professor Malachy Carey, the Associate Editor of the journal, for their constructive suggestions, which led to a substantial improvement of the paper. The first author also thanks Ed Klotz of IBM for his assistance in tackling bugs in the CPLEX toolbox for Matlab. The views are those of the authors alone.

### References


Appendix 1: Modeling variable train speed

The hypergraph-based model can be extended to allow for variable speed for a train moving between two blocks—as long as the average speed during the move equals the prescribed train speed. To this end, we can follow Sahin et al. (2008) and consider a finite number of discrete speed choices. For example, we may assume that there exist three speed options $S_1, S_2,$ and $S_3$ for train $r$. Then train $r$ departing from block $i$ at $u$ may arrive at block $j$ at $v_1 = u + \frac{j-i}{S_1}$ or $v_2 = u + \frac{j-i}{S_2}$ or $v_3 = u + \frac{j-i}{S_3}$. Correspondingly, we will have three decision variables $x^r_{i,j,u,v_1}, x^r_{i,j,u,v_2},$ and $x^r_{i,j,u,v_3}$ for this train movement, and only one out of the three variables will be selected (i.e., given value one) for a train move given $i$, $j$, and $u$. Implementation of this approach will require augmenting the sets of feasible train path arcs $\Psi^{p,r}$ and $\Psi^{f,r}$ in the preprocessing stage. It should be noted that the available speed choices will depend on the way the rail line is characterized (e.g., the distance of a block) and the length of time units. With the possibility of variable speed, constraint (7) is no longer sufficient to maintain the order of passenger subtrains. Constraint (A.1) will need to be added to the passenger train scheduling problem to control for the arrival times of passenger subtrains. This new constraint indicates that the $n^{th}$ ($\forall n = 2,3,\ldots, N$) subtrain traveling between station pair $w$ should always arrive at the destination end of $w$ no earlier than the $(n - 1)^{th}$ subtrain traveling between $w$.

\[
\sum_{(d,w,e',u',v') \in \Psi^{p,r}_{n-1}} v' \ast x^{w}_{d,w,e',u',v'} \leq \sum_{(d,w,e,u,v) \in \Psi^{p.r}} v \ast x^{w}_{d,w,e,u,v} \quad (A.1)
\]

$\forall w \in W, \forall n = 2,3,\ldots, N$
Appendix 2: The effect of allowing for en-route stops for passenger trains

Considering flying meets (i.e., no en-route stops) could constrain passenger train timetables and result in suboptimal solutions. However, given that passenger rail services considered in our study are not frequent, as is usually the case in the US, the impact of allowing for train en-route stops would likely not be significant. To confirm this speculation, we conduct a set of numerical experiments in which en-route stops of passenger trains are permitted. The objective (on the passenger side) is still minimizing total passenger schedule delay cost (later we also consider an alternative objective), with an upper bound for en-route stopping time added to the model. As mentioned in Appendix 1, constraint (A.1) is used to maintain the order of trains at the destination end of each station pair. It should be noted that overpassing is not likely for most Amtrak services and the order of passenger subtrains should be maintained throughout the way (we confirmed this through communication with staff at the Amtrak Chicago Control Center). In addition to constraint (A.1), we further introduce MAES (Maximum Allowed En-Route Stops) which gives an upper bound of time units a passenger train can spend on en-route stopping.

Table A.1 reports total passenger schedule delay cost as we incrementally increase MAES from 0 to 5 and also when MAES takes a very large number (200, which is larger than the total number of time units considered in our problem). MAES = 200 essentially describes the case that there is no constraint on how long a passenger train can stop en-route. In our sample problem (subsection 5.2), if one passenger train is dispatched in each direction, allowing for en-route stops has no influence on total passenger schedule delay cost. If more passenger trains are scheduled (2-6 in each direction), then the reduction of total passenger schedule delay cost compared to the case with no en-route stops is always less than 0.1%. In addition, we find that the values of the objective function are the same for MAES ≥ 1. Similar results are obtained for the larger problem (subsection 5.5). For example, we find that allowing each passenger train to stop en-route for up to five time units will reduce total passenger schedule delay cost only by 1.3%.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Number of passenger trains</th>
<th>Maximum allowed en-route stop time units</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Small problem (subsection 5.2)</td>
<td>1</td>
<td>179.311</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>56.110</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>50.040</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>49.094</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>44.073</td>
</tr>
<tr>
<td>Larger problem (subsection 5.5)</td>
<td>5</td>
<td>97.042</td>
</tr>
</tbody>
</table>

One may argue that once en-route stopping is permitted, the passenger-side objective function should also include en-route stopping related costs, which consists of passenger time values and train operating cost. The objective function will then be the sum of passenger schedule delay cost, passenger en-route stopping cost, and train operating cost while en-route stopping. However, directly modeling passenger en-route delay cost is difficult, because it requires tracking the number of passengers onboard on each train at any given time, and cubic terms of the decision variable $x_{i,l,u,v}$ would appear in the objective function. This imposes considerable computational challenges to solving the optimization model. On the other hand, the non-dense nature of the passenger train schedule in our study allows us to construct tight bounds for the solution of the new passenger-side optimization problem. Below we provide the details.

If we allow for en-route stopping of passenger subtrains, the new passenger-side problem can be expressed in the following simplistic form:

$$\begin{align*}
\min & \quad SD + ES \\
\text{s.t.} & \quad MAES = 200
\end{align*}$$

(A.2)

where $SD$ denotes passenger schedule delay cost; $ES$ indicates en-route stopping cost, covering both passenger time values and train operating costs related to en-route stopping.
The problem solved in Table A.1 which minimizes passenger schedule delay cost while allowing for en-route stopping can be expressed as:

\[
\begin{align*}
\text{min } SD \\
\text{s.t. } MAES &= 200
\end{align*}
\] (A.3)

The problem in the main text of our paper which minimizes passenger schedule delay cost but not allowing for en-route stopping is:

\[
\begin{align*}
\text{min } SD \\
\text{s.t. } MAES &= 0
\end{align*}
\] (A.4)

Let \( S_0 \), \( S_1 \), and \( S_2 \) denote the objective values from solving problems (A.2), (A.3), and (A.4) respectively. Note that problems (A.2) and (A.3) have the same constraints but different objective functions. Thus the solution of problem (A.2) will be a feasible solution of problem (A.3). In other words, the passenger schedule delay cost obtained from the optimal solution of problem (A.2), which is part of \( S_0 \), must be no less than \( S_1 \): \( S_0 \geq S_1 \). Between problems (A.2) and (A.4), we can view problem (A.4) as a constrained version of problem (A.2), by prohibiting en-route stops (thus en-route stopping cost \( ES = 0 \)). As a result, \( S_2 \geq S_0 \). So overall, we have \( S_2 \geq S_0 \geq S_1 \), i.e., the optimal value from minimizing passenger schedule delay cost, passenger en-route stopping cost, and train en-route stopping cost should lie between the optimal values from minimizing passenger schedule delay cost, by setting \( MAES \) to be 0 and a big number.

Table A.1 already shows that the difference between \( S_2 \) and \( S_1 \) is really small. For the sample problem, the difference between \( S_1 \) and \( S_2 \) is less than 0.1% the value of \( S_2 \); for the larger problem, \( S_1 \) is 2.2% smaller than \( S_2 \). We thus conclude that for our problems with non-dense passenger train schedules, considering flying meets improves computational efficiency while yielding results that are very close to the optimum if en-route stopping of passenger trains is allowed.
Appendix 3: Train paths for the small sample problem

Figure A.1: Train paths with two passenger trains

Figure A.2: Train paths with four passenger trains

Figure A.3: Train paths with six passenger trains

Figure A.4: Train paths with eight passenger trains

Figure A.5: Train paths with ten passenger trains

Figure A.6: Train paths with twelve passenger trains