INTEGRATED MODELING OF HIGH PERFORMANCE PASSENGER AND FREIGHT TRAIN OPERATION PLANNING ON SHARED USE RAIL CORRIDORS: A FOCUS ON THE US CONTEXT

Ahmadreza Talebian, Bo Zou

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

Abstract: This paper studies strategic level train operation planning on shared use passenger and freight rail corridors. With comprehensive consideration of realistic values for different cost components involved and the fact that passenger trains are given scheduling priority over freight trains on shared corridors in the US, we develop a hypergraph based, two-level modeling approach in which passenger and freight side costs are sequentially minimized. We explicitly consider passenger schedule delay and freight foregone demand as a function of train schedules, which are largely ignored in previous research. In particular, incorporating passenger schedule delay makes the passenger train scheduling a quadratic integer programming problem. We explore different solution approaches and conclude that a modified linearized formulation which takes advantage of the special structure of the problem achieves superior computational performance. The model is applied to a sample problem and a real-world shared use corridor in the US. We find that schedule delay cost is as important as rail fare. Scheduling more passenger trains on a shared corridor lowers passenger schedule delay but at the price of freight side cost increase. The resulting marginal freight cost increase is in most cases higher than the marginal passenger schedule delay reduction, especially when frequent passenger train services already exist on the corridor. The train speed heterogeneity significantly affects freight side cost, most of which comes from foregone demand. Having greater tolerance of degrading on-time performance of freight trains reduces foregone demand but leads to more departure and en-route train delays.

Key words: shared use rail corridor, strategic train operation planning, hypergraph, integer programming, passenger schedule delay, foregone freight demand.

1 Introduction

Passenger rail has been resurging in the United States over the past decade. Amtrak, the primary intercity rail service provider in the US, has witnessed continuous ridership growth (except for 2009 due to the economic recession) from 20.9 million to 31.6 million passengers, or 51.1% increase, between 2000 and 2013 (Amtrak, 2013). To sustain this trend which promotes sustainability and multimodality at the same time in inter-city travel, several states have been pursuing higher performance rail systems, in the forms of High Speed Rail (HSR) services on new or existing corridors. While California has chosen to build a brand-new, dedicated HSR line which costs about $68 billion dollars (California HSR, 2012), a more progressive approach is adopted in the Midwest Region. A prominent example is the Chicago—St. Louis corridor, where the existing single-track line is being upgraded to accommodate passenger trains running at a maximum speed of 110 mph. Once this track upgrade phase is complete in 2017, total line-haul travel time will be reduced by one hour (Illinois HSR, 2014). The ensuing phase pertains to adding more sidings along the line to increase capacity. By taking advantage of the existing rail infrastructure, the construction cost is much lower than in the California case, in the magnitude of a few billion dollars for the initial phase (Illinois HSR, 2014). The higher speed trains often share the use of tracks with freight trains, and the tracks are, in most cases, owned and maintained by freight railroads (Peterman et al., 2009). The mixed operations give rise to the issue of capacity constraints, as manifested by the non-trivial delays reported by Amtrak and freight trains on shared use corridors (Harrod, 2009). It is therefore important to understand the interactions between passenger and freight operations from

---

1 Corresponding author.

Email addresses: ataleb2@uic.edu (A. Talebian), bzou@uic.edu (B. Zou).
the strategic level scheduling perspective, especially given the introduction of more high performance passenger rail services on shared use corridors in the future.

To describe the "strategic level scheduling", an understanding of the train planning process is necessary. On the passenger side, a typical rail planning process consists of six stages, according to Ghoseiri et al. (2004). In the first stage, rail passenger demand for each Origin-Destination (OD) pair is determined, accompanied sometimes by a sketch of the temporal distribution of demand in coarse periods of time (e.g. every 15 minutes). Then comes the second stage, line and frequency planning, which decide on what routes to serve, and the frequency of service. Given the train frequency on a route, the schedules (timetables) of trains are specified at the third stage. The fourth stage deals with rolling stock planning in which locomotives and cars are assigned to trains. This is followed by crew scheduling and rostering (stages 5 and 6), which attempts to assign crews to trains, and construct rosters from the crew duties. Ghoseiri et al. (2004) further propose to group the six stages into two planning levels: strategic and tactical, with train scheduling lying on the border between the two levels. In light of these, in this paper we define strategic level passenger rail scheduling as the activity of determining preliminary, non-minute-by-minute, passenger train timetables given passenger demand and train service frequency.

On the freight side, train scheduling process is quite different in North America: freight train schedules are determined much closer to the actual train departure time than passenger train schedules—sometimes only one day before. It is also possible that a freight train runs without a prescribed schedule (Mu and Dessouky, 2011). In day-to-day operations, freight train departures are more a function of demand: a train simply departs once it receives sufficient tonnage of load (Cordeau et al., 1998). Given the less precise and stringent nature of freight train schedules, the "strategic level scheduling" refers to modeling freight train operations that takes account of overall freight demand and consequently service frequency of trains, without going into detailed operating characteristics.

Development of travel surveys coupled with advanced behavioral modeling techniques has made possible detailed inference about passenger travel demand preferences. This is further accompanied by the emergence and implementation of automated boarding and alighting counting systems to collect passenger origin-destination demand matrices. These advancements call for new train scheduling models taking advantage of detailed passenger information, in which passengers’ preferred time of travel is of particular interest. For scheduled transportation services, passenger preferred time of travel directly relates to the concept of schedule delay, particularly when services are relatively infrequent (Kanafani, 1983). For a traveler, schedule delay is defined as the difference between one's desired departure time and the actual departure time (Hendrickson and Kocur, 1981). The time difference characterizes the inconvenience of schedule to accommodate one's desired activities, which incurs time-related cost to the traveler. While schedule delay has been considered in air transportation (e.g. Brueckner, 2004; Brueckner and Flores-Fillol, 2007; Panzar, 1979; Richard, 2003; Schipper et al., 1998; Zou and Hansen, 2012) and transit (e.g. Alfa and Chen, 1995; Kraus and Yoshida, 2002; Tian et al., 2007), very limited attention has been paid in passenger rail service planning. Kanafani (1983) shows that for short-haul intercity air travel, schedule delay accounts for a significant portion of the total generalized cost for a traveler. Given that train service may be less frequent than flights operating on an intercity corridor, we expect an even more important role schedule delay will play in strategic level passenger rail scheduling.

Quantifying passenger schedule delay requires knowledge about not only train departures but also the "ideal" departure time of each traveler, the latter of which is difficult for researchers or planners to estimate with precision. In the state-of-the-practice intercity travel demand forecasting, surveys are conducted to construct passenger preferred departure time (PDT) profiles, in which PDTs are discretized into relatively coarse time intervals (e.g. 15-min intervals in Cascetta and Coppola (2012)). Because of this discretization, it may be more sensible, from the strategic level scheduling perspective, to produce train schedules with similar time resolutions rather than minute-by-minute timetables.

The objective of the present study is therefore to introduce the concept of strategic level scheduling to train operational planning on shared use passenger and freight rail corridors. The strategic level scheduling attempts to produce train timetables that are in accordance with the operational nature of freight train operations and also consideration of passenger schedule delay, the latter is an integral component in intercity passenger transportation but often ignored in existing train scheduling research and practice. Therefore, incorporating passenger schedule delay represents one important contribution of this research. To model train operations and produce optimal timetables, a directed hypergraph based approach is employed, which is introduced to rail scheduling only recently and has so far seen very limited applications (Harrod, 2009; 2011). As we will
show later, the directed hypergraph based approach is superior to conventional models by addressing an important omission of train path conflicts due to track resource use during transition.

Our study also differs from previous research by considering a comprehensive set of cost components involved in passenger and freight train operations. We further recognize that, by the US Federal law (110 Congress, 2008; Harrod, 2009; Wilner, 2013), passenger trains are given scheduling priority on shared use rail corridors. In light of this, we put forward a hypergraph-based, two-level nonlinear integer programming framework, wherein passenger trains are first scheduled, after which freight trains are inserted among passenger trains. In particular, we develop schedules for a single type of passenger trains, i.e., all passenger trains run at the same speed and stop at the same stations. On the other hand, freight trains are flexible: they can travel between any station pair, stop within the route, and run at different speeds. We explore multiple solution approaches aiming at tackling computational complexities of the formulation and identify the best strategy for solving the strategic level problem of train scheduling on shared use corridors. Our model reflects the nature of elastic passenger demand with respect to train service frequency, and foregone demand which occurs to freight operators due to line capacity constraints. We perform an extensive examination of the impact on system performance of passenger train service frequencies, speed heterogeneity between passenger and freight trains, and the tolerance of worst-scenario freight train on-time performance. They are all related to strategic decisions and therefore consistent with our concern of strategic level scheduling. We use realistic parameter values, which are not commonly seen in theoretic rail scheduling studies, and demonstrate an application of the strategic level scheduling model to the Chicago-St Louis higher speed rail line. The results obtained from our modeling and analysis provide not only inputs for subsequent, shorter-term train schedule planning at the tactical and/or operational levels, but also practical and general insights into future rail planning on shared use corridors.

The remainder of the paper is organized as follows. In Section 2, a review of train scheduling literature is performed, based on which we highlight the contribution of the present research. Mathematical formulation of strategic level train scheduling problem is presented in Section 3, followed by discussions on the solution strategies in Section 4. Section 5 performs numerical analyses, on both a small sample problem and the Chicago-St Louis HSR case study. Sensitivity analysis on speed heterogeneity and freight train delay tolerance are also conducted. Section 6 summarizes the major findings from the paper and offers directions for future research.

2 Literature review and research contribution

Train scheduling research can date back to the 1960s, when mathematical models are first developed to address two-way traffic on a single rail line (Frank, 1966). Since then various approaches have been developed to tackle different issues facing train scheduling, and in general they can be classified into three categories: analytical, simulation, and discrete optimization approaches (Abril et al., 2008). The analytical approach, especially popular in the early days of train scheduling research, uses simple models to estimate the capacity of a rail line, 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). Despite its simplicity, the applications of analytical models are constrained in the real world as they are only able to capture train operational characteristics to a limited degree. As a result of this and also thanks to significant advances in computational power, recent research has mostly shifted to simulation and optimization approaches, which offer greater flexibility to model details of train scheduling problems.

In rail industry practice, simulation is the dominant method for train scheduling. Popular software such as Rail Traffic Controller (RTC) (Willson, 2012) and Módulo Optimizador de Mallas (MOM) (Barber et al., 2006) incorporate various parameters, including train types, equipment types, terrain and track conditions, train speed, acceleration and deceleration, and traffic signals, to simulate real world train dispatching and operation. However, passenger side cost, to our knowledge, has not yet been explicitly incorporated in the existing simulation tools. In addition, the simulation approach does not address train scheduling in a system optimum way, but attempts to emulate human dispatcher behavior in resolving train conflicts (Harrod, 2013). Cognizant of the weakness of the simulation approach in improving system performance, Jovanović and Harker (1991) combine simulation and optimization, and employ a variant of branch-and-bound partial enumeration procedure to generate robust schedules at the tactical level, which satisfy the physical constraints of over-the-line train operations. However, such attempt is rare, not only due to the complexity of developing hybrid models, but also because of the commercial nature of most simulation tools, which limits the access of rail operations researchers to the inner core of the tools.
In contrast to the simulation approach, the discrete optimization approach is able to identify train schedules that generate system optimum performance, which is of particular value in strategic rail planning. Discrete optimization has been used for train scheduling since Amit and Goldfarb (1971) who introduce the first mathematical model to train timetabling. Over the past decades, numerous studies have appeared in the literature. In what follows we only review some of the most recent and relevant ones. A summary of these studies is presented in Table 1.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Objective</th>
<th>Modeling priority</th>
<th>Discrete time</th>
<th>Model structure</th>
<th>Solution method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ceder (1991)</td>
<td>Min maximum headway</td>
<td>N</td>
<td>Y</td>
<td>ILP</td>
<td>Heuristic based on the shortest path algorithm</td>
</tr>
<tr>
<td>Nachtigall (1996)</td>
<td>Min total waiting time at stations</td>
<td>N</td>
<td>N</td>
<td>ILP</td>
<td>Branch-and-bound</td>
</tr>
<tr>
<td>Brännlund et al. (1998)</td>
<td>Min schedule deviation</td>
<td>Y</td>
<td>Y</td>
<td>ILP</td>
<td>Lagrangian relaxation</td>
</tr>
<tr>
<td>Oliveira and Smith (2000)</td>
<td>Min schedule deviation</td>
<td>N</td>
<td>Y</td>
<td>ILP</td>
<td>Constraint programming</td>
</tr>
<tr>
<td>Caprara et al. (2002)</td>
<td>Min schedule deviation</td>
<td>N</td>
<td>Y</td>
<td>ILP</td>
<td>Lagrangian relaxation</td>
</tr>
<tr>
<td>Ghoseiri et al. (2004)</td>
<td>Min total passenger-time and fuel consumption cost</td>
<td>N</td>
<td>N</td>
<td>MINLP</td>
<td>Pareto optimality</td>
</tr>
<tr>
<td>Caprara et al. (2006)</td>
<td>Min schedule deviation</td>
<td>Y</td>
<td>Y</td>
<td>ILP</td>
<td>Lagrangian relaxation</td>
</tr>
<tr>
<td>Yang et al. (2009)</td>
<td>Min total passenger travel time and total delay</td>
<td>N</td>
<td>N</td>
<td>MILP</td>
<td>Fuzzy simulation-based</td>
</tr>
<tr>
<td>Caprara et al. (2010)</td>
<td>Min schedule deviation</td>
<td>Y</td>
<td>Y</td>
<td>ILP</td>
<td>Branch-and-bound</td>
</tr>
<tr>
<td>Harrod (2011)</td>
<td>Max total utility of trains</td>
<td>Y</td>
<td>Y</td>
<td>ILP</td>
<td>Branch-and-bound</td>
</tr>
<tr>
<td>Liu and Kozan (2011)</td>
<td>Min schedule makespan</td>
<td>Y</td>
<td>N</td>
<td>MILP</td>
<td>Hybrid heuristic</td>
</tr>
<tr>
<td>Canca et al. (2012)</td>
<td>Minimum total arrival time of special service trains</td>
<td>N</td>
<td>Y</td>
<td>ILP</td>
<td>Branch-and-bound</td>
</tr>
<tr>
<td>Li et al. (2013)</td>
<td>Min energy and carbon emission cost and total passenger travel time</td>
<td>N</td>
<td>N</td>
<td>MILP</td>
<td>Fuzzy mathematical programming</td>
</tr>
<tr>
<td>Barrena et al. (2014)</td>
<td>Min passenger waiting time</td>
<td>N</td>
<td>Y</td>
<td>MILP</td>
<td>Branch-and-cut</td>
</tr>
<tr>
<td>Canca et al. (2014b)</td>
<td>Min passenger waiting time</td>
<td>N</td>
<td>Y</td>
<td>INLP</td>
<td>NLP-based branch-and-bound</td>
</tr>
</tbody>
</table>

Note: INLP: Integer Non-Linear Programming; ILP: Integer Linear Programming; MILP: Mixed Integer Linear Programming; MINLP: Mixed Integer Non-Linear Programming.

Most train scheduling problems have been modeled as discrete time networks with multi-commodity flows through constrained resources. Brännlund et al. (1998) present a binary linear programming for scheduling passenger and freight trains on a single track corridor. Assuming that each train receives a certain profit when following a given timetable, the model maximizes the sum of profits of all trains subject to constraints on track capacity. Using a Lagrangian relaxation approach, the authors separate the original problem into one independent shortest path subproblem for each physical train. Caprara et al. (2002; 2006) find the best actual train timetables that have the smallest deviation from the ideal timetables in a one-way track setting. An ideal timetable represents the most desirable timetable for a given train. This actual timetable is modifiable to meet the track capacity constraint. The problem is formulated as a binary linear program that maximizes the sum of profit across arcs, which is shown equivalent to minimizing schedule deviation, subject to track capacity and flow conservation constraints. Caprara et al. (2010) further extend Caprara et al.’s models to take account of bi-directional train traffic, arbitrary network topology, and train rerouting. In their model, passenger trains stick to the prescribed timetables, whereas freight trains receive ideal timetables which may be modified by the infrastructure manager. The objective function minimizes the sum of deviations from the ideal timetable due to departure time shifting, stopping, and rerouting.
The concept of job-shop scheduling has also been applied to train scheduling research. Oliveira and Smith (2000) consider train trips as a set of jobs which are scheduled on resources (tracks). Total train departure delays are minimized with respect to predetermined ideal train timetables, subject to meeting requirements between two opposing trains, minimum headway between two consecutive trains, and potential blocking of the tracks. Liu and Kozan (2011) propose a job-shop scheduling based method for scheduling priority passenger and non-priority freight trains on a single track railroad. Makespan is minimized subject to time consistency, priorities, and blocking constraints. In their model, minimizing makespan can be translated to minimizing total length of train schedule. As a heuristic algorithm is employed to solve the job-shop scheduling problem, the optimality of the solution is not guaranteed. A similar idea is introduced by Zhou and Zhong (2005) where a multi-mode flow-shop scheduling model is put forward to schedule high speed and medium speed trains on a double track line. Different from other studies, the authors model train acceleration and deceleration by considering multiple execution modes for a train traversing a rail section. They consider two objective functions: variation of interdeparture times for high speed trains and total travel time of all trains. Arguing that the conventional sequential scheduling approach degrades the service quality of medium speed trains, the authors use a beam search algorithm to generate non-dominated schedules guaranteeing appropriate service quality for all trains.

We note that most of the existing studies ignore the effects of train schedules on passengers. Only a few attempts have been made to investigate the trade-off between passenger total travel time and train operation performance. Ghoseiri et al. (2004) develop a multi-objective optimization model to schedule passenger trains on single or multiple track railways. Fuel consumption and total passenger time are used to represent the operator and users interests. A Pareto frontier is determined at the first stage of a two-stage solution procedure. In the second stage, the authors utilize the obtained frontier in a distance-based multi-objective optimization to find the schedule that maximizes the weighted sum of normalized objectives, i.e. fuel consumption and total passenger time. A mixed integer programming formulation is proposed by Li et al. (2013) to minimize energy and carbon emission cost, and the total passenger travel time. A fuzzy multi-objective optimization algorithm is developed to find a timetable simultaneously minimizing both objectives. A goal programming approach is employed in Yang et al. (2009) to solve train scheduling problems. Arguing that the real-world systems always work under uncertainty, the authors consider minimizing the fuzzy total of passenger in-vehicle time and total delay at stations. The optimal solution is obtained by using a branch-and-bound algorithm based on fuzzy simulation.

As is mentioned in Section 1, an important component in passenger travel time is schedule delay, which is only tangentially touched upon in limited studies. Ceder (1991) presents a model which reduces passenger waiting time at train stations while maintaining the minimum number of trains required by the operator. Nachtigall (1996) sets a boundary for each running and stopping activity of trains and finds periodic timetables through minimizing weighted sum of passenger waiting times at stations. These studies, however, do not account for the temporal distribution of passengers arriving at the stations. Canca et al. (2012) employ the dynamic behavior of demand to find optimal rescheduling policies, including short-turning and deadheading. They suppose that travel demand between a specific station pair increases to an extent to which a special train service, named shuttle, is required to serve demand. The objective is therefore to insert the optimal number of shuttles, which do not stop at stations with low passenger demand, into an existing schedule of trains. As they assume passenger demand between each station pair is uniformly distributed over the course of the day, the resulting schedule minimizes total passenger waiting time. Of course, the assumption of uniform passenger demand distribution is ideal and needs to be adjusted when implementing the model.

The most relevant studies to the problem studied in our paper are Canca et al. (2011; 2014a), which propose a non-linear integer programming approach to tackle passenger train scheduling on a double track rail line. Passenger demand functions for each OD pair are constructed, with the objective of minimizing passenger average waiting time. However, the sensitivity of passenger demand to train frequency is not recognized. In a similar study, Barrena et al. (2014) propose multiple linearized formulations to minimize total passenger waiting time at stations. Canca et al. (2014b) extend the work of Canca et al. (2014a) to take into account the elasticity of passenger demand with respect to service frequency. To this end, the authors assume that long waiting times causes travelers to shift to an alternative mode. The problem is formulated as minimization of the passenger loss probability. They propose a logit-based model and a sigmoid function-based model to calculate the probability. In particular, the logit-based model takes into consideration the impacts on passenger demand of services frequency change. The three aforementioned studies merely focus on passenger
train traffic and lack system-wide cost consideration. In addition, modeling is restrained to one-way, single-type train traffic.

Whereas most of the existing studies under the discrete optimization category formulate time-block occupancy programs to generate feasible train timetables, an alternative, hypergraph based approach is recently introduced to train scheduling (Harrod, 2009; 2011). Compared to the conventional discrete time dynamic graphs and block occupancy conditions which are frequently used to model multi-commodity flows or activity paths through constrained resources, Harrod demonstrates that the hypergraph based formulation addresses a critical omission in the traditional discrete time-block occupancy programs of train transitions between two consecutive blocks, which can lead to undetected train paths conflicts.

Overall, in spite of the rich body of existing train scheduling studies, significant gaps remain in the literature. First, there seems a lack of research on strategic level train scheduling on shared use passenger and freight corridors when one type of trains is given scheduling priority. In this case, the interaction of different train types is not well understood. Second, we incorporate schedule delay, a proxy of inconvenience of schedule to travelers, into the train scheduling problem using passenger preferred departure time distribution. Third, we explore several approaches aiming at tackling computational complexities of the quadratic binary programming problem and introduce a formulation capable of solving real-world size problems within reasonable times. Fourth, existing train scheduling models often use presumed and ad hoc parameter values, and only a part of the systematic cost components is considered. In particular, costs associated with passenger schedule delay and foregone freight demand are important in designing optimal train schedules but are largely absent in existing studies. These significantly compromise the power of the developed models to support real-world rail planning and decision making. Fifth, we analyze the impacts on freight related costs of maximum delay tolerance level, a strategic decision which is not sufficiently examined. Sixth, while binary integer occupancy programming is the prevailing choice for modeling, the emerging hypergraph based approach in rail scheduling seems superior, yet deserves further investigation, in both model formulation and computation. This is especially important with systematic consideration of different stakeholders’ interests on shared use passenger and freight corridors. These gaps will be filled in the present research.

3 The model

In this section we propose a sequential modeling approach to scheduling passenger and freight trains on a share use corridor. We first introduce the hypergraph concept for characterizing train movements, after which the train scheduling models are specified. The sets, parameters, and decision variables used in our models are documented in Table 2.

<table>
<thead>
<tr>
<th>Type</th>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary decision</td>
<td>$x_{i,j,u,v}$</td>
<td>Occupancy arc representing the possession of node $i$ at time $u$ and the exit into node $j$ at time $v$ of subtrain $r$</td>
</tr>
<tr>
<td></td>
<td>$y_{r,t}^f$</td>
<td>Artificial arc linking arrival of subtrain $r$ at a station at time $t$ to departure of its continuation subtrain $f$ at time $t$</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 opposite 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 opposite 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 all track blocks, ordered by a common reference of travel such as “North” or “South”</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 sharing equipment resources (i.e., the same physical train)</td>
</tr>
<tr>
<td>Type</td>
<td>Component</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>( \Psi_{P,r} )</td>
<td>( (i,j,u,v) )</td>
<td>The set of feasible arcs ((i,j,u,v)) for passenger subtrain ( r ) supplied from preprocessing</td>
</tr>
<tr>
<td>( \Psi_{T,r} )</td>
<td>( (i,j,u,v) )</td>
<td>The set of feasible arcs ((i,j,u,v)) for freight train ( r ) supplied from preprocessing</td>
</tr>
<tr>
<td>( T )</td>
<td></td>
<td>The set of network cells</td>
</tr>
<tr>
<td>( L_{r}^P )</td>
<td></td>
<td>The set of valid pairs of arrival times for passenger subtrain ( r ) and departure of its continuation ( \hat{r} ): ((d^w,e^r,u,t) \in \Psi_{P,r}, (o^w,j,\hat{t},v) \in \Psi_{P,r}, t + l_{\min}^r \leq \hat{t} \leq t + l_{\max}^r)</td>
</tr>
<tr>
<td>( W )</td>
<td></td>
<td>Set of station pairs</td>
</tr>
</tbody>
</table>

### Parameters

- \( o^w \) Origin block of station pair \( w \)
- \( d^w \) Destination block of station pair \( w \)
- \( EADT^r \) Earliest allowed departure time (EADT) from origin of train (or subtrain) \( r \)
- \( LAAT^r \) Latest allowed arrival time (LAAT) at destination of train (or subtrain) \( r \)
- \( l_{\text{max}}^r \) Maximum allowable layover time at a station for passenger subtrain \( r \)
- \( l_{\text{min}}^r \) Minimum allowable layover time at a station for passenger subtrain \( r \)
- \( c_i^r \) Cost per unit time of layover for passenger subtrain \( r \)
- \( c_p^r \) Foregone demand cost for freight train \( r \)
- \( c_i^r \) Operating cost of stopping status plus freight value of time per time unit for freight train \( r \)
- \( c_m^r \) Operating cost of moving status plus freight value of time per time unit for freight train \( r \)
- \( b_i^t \) Capacity (count of trains) of block \( i \) at time \( t \)
- \( v_i^t \) Capacity (count of trains) of cell \( i \) at time \( t \)
- \( \varepsilon \) Leading transition time margin
- \( \delta \) Lagging transition time margin
- \( h^r \) Minimum gap between train (or subtrain) \( r \) and following trains
- \( c_d^w \) Value of schedule delay time for travelers whose PDT is before a train departure, measured in $/hour
- \( c_d^w \) Value of schedule delay time for travelers whose PDT is after a train departure, measured in $/hour
- \( q_{m}^w \) 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 farther than the destination of station pair \( w \)), and desire to leave between \( t=z-1 \) and \( t=z \)

### Sink node \( e^r \)

- Artificial sink node designating that train (or subtrain) \( r \) is off the network

### Intermediate variables

- \( C_{u,L}^{r_N^w} \) Total schedule delay cost of passengers who prefer to depart between \( \hat{u} \) and \( \hat{u} \) and end up boarding subtrain \( r_N^w \)
- \( C_{u,R}^{r_N^w} \) Total schedule delay cost of passengers who take subtrain \( r_N^w \) and prefer to depart between \( \hat{u} \) and \( \hat{u} \), where \( \hat{u} \) is a given departure time of subtrain \( r_N^w \) and end up boarding subtrain \( r_N^w \)
- \( C_{u,R}^{r_N^w} \) Total schedule delay cost of passengers who take subtrain \( r_N^w \) and prefer to depart between \( \hat{u} \) and \( \hat{u} \), where \( \hat{u} \) is a given departure time of subtrain \( r_N^w \) and end up boarding subtrain \( r_N^w \)
- \( C_{u,L}^{r_N^w} \) Total schedule delay cost of passengers who take subtrain \( r_N^w \) and prefer to depart between \( \hat{u} \) and \( \hat{u} \), where \( \hat{u} \) is a given departure time of subtrain \( r_N^w \) and end up boarding subtrain \( r_N^w \)
- \( C_{u,R}^{r_N^w} \) Total schedule delay cost of passengers who take subtrain \( r_N^w \) and prefer to depart between \( \hat{u} \) and \( \hat{u} \), where \( \hat{u} \) is a given departure time of subtrain \( r_N^w \) and end up boarding subtrain \( r_N^w \)
- \( C_{u,L}^{r_N^w} \) Total schedule delay cost of passengers who take subtrain \( r_N^w \) and prefer to depart between \( \hat{u} \) and \( \hat{u} \), where \( \hat{u} \) is a given departure time of subtrain \( r_N^w \) and end up boarding subtrain \( r_N^w \)
### Type | Component | Description
--- | --- | ---
$C_{u,L}^r$ | Total schedule delay cost of passengers who prefer to depart before $u$ (departure time of subtrain $r_w^v$) and end up boarding subtrain $r_w^v$ |  
$C_{u,R}^r$ | Total schedule delay cost of passengers who prefer to depart between $u$ (departure time of subtrain $r_w^v$) and $t=T$ and end up boarding subtrain $r_w^v$ |  

### 3.1 Hypergraph concept

As mentioned before, discrete time-block dynamic graphs are widely used to model multi-commodity flows in solving train scheduling problems. However, this approach is found incapable of capturing the unintended violation of train paths during transitions between blocks (Harrod, 2011). The left panel of Figure 1 briefly illustrates this. The horizontal axis denotes discrete time periods, and the vertical axis represents tracks segment blocks. Assuming a single-track line, the paths of two opposing trains shown do not violate block occupancy constraints, as each single block is only occupied by one train at any time period. However, the train transition that occurs at the end of the $(t + 1)^{th}$ period on the boundary of blocks 3 and 4 is clearly infeasible, because there is no additional track to allow the two trains to meet and pass. This path conflict is omitted when using the traditional discrete time-block dynamic graph.

![Figure 1: Illustration of unintended train path conflict using conventional discrete time dynamic graph and hypergraph (adapted from Harrod (2011))](image)

To account for the potential train path conflicts during transitions, we employ a hypergraph-based train scheduling model. A hypergraph is a generalized graph which allows an edge (or arc) to connect multiple nodes. A hypergraph can be uniquely defined by $\mathcal{H} = (V, E)$, where $V$ is the set of vertices (nodes) and $E$ is the set of hyperarcs (hyperedges), of which each hyperarc $e \in E$ maps to a non-empty set of vertices drawn from $V$. In this paper, we follow Harrod (2011) and use directed hypergraphs to model train moves on rail tracks. An example of a hypergraph is shown in the right-hand-side graph of Figure 1. This hypergraph, which is essentially a time expansion of railway tracks, entails two types of vertices: block occupancy nodes (solid points) and cells (hollow points). The cells intend to capture train transition from one block to another block. Each block occupancy node is labeled by its track number and the corresponding occupation time: $(i, t) \in \{(i, t) | i \in B, t \in T\}$. Each cell is labeled by its lower left block occupancy node. In the right panel of Figure 1, the outlined cell is labeled $(3, t)$.

Each movement of a train is described by a hyperarc, an example of which is shown in the right panel of Figure 1 as circled. This hyperarc consists of two block occupancy nodes $(1, t)$ and $(2, t + 1)$, and one cell $(1, t)$, denoting that the train occupies block 1 at period $t$, transitions through cell $(1, t)$, and occupies block 2 at period $t + 1$. An unambiguous indicator of this hyperarc must entail time- and space-related characteristics of tail and head nodes, as well as the train index. Therefore, a hyperarc can be recognized by a 0-1 indicator $x_{i,j,u,v}^r$ which
denotes that train \( r \) occupies node \( i \) at period \( u \) and node \( j \) at period \( v \). For the circled hyperarc in Figure 1, the associated 0-1 indicator is \( x_{u,2,3,4,t+1}^i \). Obviously, there is no local transition if \( i = j \), and the associated hyperarc is horizontally positioned. To identify train path conflicts, we compare the hyperarcs of train paths and determine whether the nodes and cells covered by the active hyperarcs in each time period are used by more trains than the capacity of the nodes and cells. For example, in the left panel of Figure 1 the movement of the black train \( (train \ b) \) from block 3 to block 4 can be denoted by hyperarc \( x_{3,4,t+1,1,t+2}^b \); the movement of the grey train \( (train \ g) \) from block 4 to block 3 is expressed using hyperarc \( x_{4,3,t+1,1,t+2}^g \). These two hyperarcs both contain cell \((3, t+1)\) with the capacity of one. Therefore, the train path conflict during transition is detected.

In a hypergraph, a chain of consecutive hyperarcs form the train path. In Figure 2, the path of the gray train \( g \) is characterized by a series of hyperarcs: \( x_{2,4,3,t+1}^g \), \( x_{4,3,t+1,1,t+2}^g \), \( x_{3,3,t+2,2,t+3}^g \), and \( x_{3,3,t+3,t+4}^g \). These hyperarcs may take one or multiple time periods. Although \( u \) and \( v \) in the above hyperarcs \( x_{u,v,t} \) are restricted to be either identical or two neighboring nodes, one could alternatively represent a complete train path using a single hyperarc, as is shown by the black train in Figure 2. In this case, the hyperarc train path covers seven block occupancy nodes and four cells. The use of more complicated hyperarcs can still capture train transitional conflicts and block occupancy limits, but would require different preprocessing to generate feasible train paths.

In the present paper, we consider the simpler version of hyperarcs and assume that each hyperarc entails a head node, a tail node, and at most one transitional node.

![Figure 2: Two representations of train paths on hypergraphs](image-url)

Before delving into decision variables in the hypergraph based model, we would like to introduce the definition of subtrains, each of which represents a subjourney of a train between two consecutive stopping stations. Therefore, the journey of a physical train traversing at least one intermediate station is modeled as a set of subjourneys, each conducted by a subtrain. The first subtrain is recognized as an unlinked subtrain, and the other subtrains each linked to its previous subtrain are named linked subtrains. Clearly, if a train’s journey does not include an intermediate station, distinguishing a subtrain and a train is not necessary. This also applies

---

2 Alternatively, one can employ a non-hypergraph-based approach, such as the one in Brännlund et al (1998), to account for train path conflicts during transitions between blocks in an implicit way. In this paper, however, we are interested in explicitly modeling the transitional status.

3 We thank one of the reviewers for the suggestion. Indeed, consideration of such more complicated hyperarcs is part of our on-going work.
to freight trains. To unambiguously express one subtrain of a physical train, we introduce \( r^w_n \) \( (n = 1, \ldots, N; w \in W) \), which denotes the \( n^{th} \) subtrain travelling from the origin block of station pair \( w \) to the destination block of station pair \( w \). Figure 3 illustrates three physical trains traversing a line, each of which is composed of three subtrains. In this example, \( N = 3 \) and \( W = \{w_1, w_2, w_3\} \). \( r^w_1 \) represents the first subtrain traveling from the origin of station pair \( w_1 \) (station 1) to the destination of station pair \( w_1 \) (station 2). Similarly, \( r^w_2 \) is the second subtrain traveling between station pair \( w_1 \). Subtrain \( r^w_3 \), which is the first subtrain traveling between station pair \( w_2 \) (i.e., from station 2 to station 3), will be the continuation of \( r^w_1 \). Based on the above definition, \( r^w_1 \) is an unlinked subtrain; whereas \( r^w_2 \) is a linked subtrain. In this study, we assume that each physical passenger train running in a given direction will stop at the same stations along the line.

Using hypergraph, finding train schedules translates into the problem of identifying the hyperarcs for each train that constitute the train’s path. As mentioned, each hyperarc is equivalent to a binary decision variable \( x^r_{i,j,u,v} \) which equals 1 if subtrain \( r \) occupies block \( i \) in time interval \([u, v)\) and moves to block \( j \in \{B|j \neq e_r\} \) at \( v \), where \( B \) denotes the set of all track blocks; and \( e_r \), the artificial sink node (which we discuss below), and 0 otherwise. In \( x^r_{i,j,u,v} \) can be either \( i \) (i.e., stopping hyperarc) or the indicator of the very next block, i.e., \( i + 1 \) or \( i - 1 \), depending on subtrain \( r \)’s moving direction. Taking into consideration the above discussion, we characterize the journey of a train as follows. Subtrain \( r \), which travels between station pair \( w \), starts its journey from the origin block \( o^w \) and arrives at the destination node \( d^w \). We further name any arc that initiates subtrain \( r \)’s journey the “starting arc”, \( x^r_{o^w,i,j,u,v} \). When subtrain \( r \) arrives at its destination \( d^w \), the subtrain sinks into an artificial sink node \( e^r \), which removes the subtrain from the network at the completion of its journey. In essence, each sinking arc, i.e., the arc connecting \( d^w \) to the sink node, is an artificial arc, which is assumed to take one unit of time. At the stopping station, another artificial arc, connecting the sink node back to the station, is also introduced to link subtrain \( r \) to a new subtrain \( r \) which starts its journey from the origin block of station pair \( \tilde{w} \). The linkage between each pair of consecutive subtrains (\( r \) and \( r \)) is established using a binary decision variable \( y_{i,j}^{r,r} \), which equals 1 if subtrain \( r \) arrives at an artificial sink node \( e^r \) at \( t \) and its continuation \( r \) resumes the journey from the origin node \( o^w \) at time \( \tilde{t} \).

To further clarify the above discussion, a physical train composed of two subtrains (\( r \) and \( r \)) is illustrated in Figure 4. Unlinked subtrain \( r \) starts its journey from block 1 (Station 1) at time period \( t \) and arrives at block 3 (Station 2) at time period \( t+2 \). The first grey hyperarc is the artificial sinking hyperarc taking one time period. The second grey arc is artificial arc \( y_{t+3,t+4}^{r,r} \) denoting the linkage between trains \( r \) and \( r \). Linked subtrain \( r \) departs its origin at time period \( t+4 \) and finishes its journey at time period \( t+7 \).
Given the pre-specified upper and lower bounds on layover (i.e., stopping at stations, for passenger boarding and alighting) times $l_{\text{max}}$ and $l_{\text{min}}$ at the preprocessing stage, we generate $y$ variables guaranteeing layover times between $l_{\text{max}}$ and $l_{\text{min}}$. Each subtrain is assigned an earliest allowed departure time ($EADT^r$) from its origin block and a latest allowed arrival time ($LAAT^r$) at its destination block. The ensuing sub-section will discuss about how $EADT$ and $LAAT$ are determined. $EADT$ and $LAAT$ information will then be used at the preprocessing stage to generate feasible hyperarcs, i.e., $x$ variable values.

### 3.2 Model formulation

In this paper, the strategic level train scheduling problem is considered from a central planner's perspective. In the US, passenger trains, including the proposed higher speed rail services, are given scheduling priority by the Federal law. To reflect this, a two-level sequential modeling approach is adopted. At the upper level, we first find passenger train schedules that minimize passenger side cost. Given the passenger train schedules, at the lower level we develop freight train timetables that minimize freight side cost. We provide detailed model formulation in this sub-section.

#### 3.2.1 Upper Level: passenger train scheduling

At the upper level, the objective of the central planner is to minimize total passenger side cost, which consists of: 1) passenger schedule delay cost; 2) passenger in-vehicle travel time cost; and 3) train operating expenses. Train ticket fare is an internal transfer between passengers and the passenger rail agency and therefore is not counted. In this study, we intend to design a schedule which permits two opposing passenger trains to pass without any full stop, which is called “flying meet” (Dure, 1999; Petersen and Taylor, 1987). Therefore, passenger in-vehicle travel time and train operating cost are invariant to train schedules. The objective function of the upper level optimization only needs to include passenger schedule delay cost.

We assume that the discrete passenger Preferred Departure Time (PDT) profile is known a priori for each direction at each train station, given the number of running passenger trains. Figure 5 illustrates passenger PDT at a station with three train departures. In the figure, each blue bar indicates the number of travelers whose PDT falls within a time interval. Each passenger departing from the station tries to minimize her schedule delay. This results in a deviation of her actual departure time from her PDT, either departing earlier or later, which is the schedule delay for the traveler. If the traveler values equally schedule displacement in either direction (earliness or lateness), her decision is simply to choose the train with the closest departure to her PDT. On the other hand, a traveler may value the inconvenience of earlier/later schedule displacement differently. An

---

4 This is done by specifying the set for feasible path arcs $\Psi^r$ (refer to Table 2), in which no stopping arc for passenger subtrains is generated.
example can be that travelers value highly being at the destination in time, and consequently penalize to a greater extent if the schedule delay is associated with taking a later train than one’s PDT. Therefore, it may be sensible to assign a larger schedule delay penalty for lateness than earliness schedule displacement. Through this passenger self-selection process of trains, we can calculate the number of passengers boarding each train and total passenger schedule delay for each train. Note that passengers also care about their Preferred Arrival Time (PAT). As we assume no en-route delay for passenger trains, one can shift every bar in the PAT distribution for each OD pair by the same in-vehicle travel time, which results in the PDT distribution.

Figure 5: Passenger PDT distribution at a train station (going one way) and self-selection of passenger trains

Special attention should be paid to setting EADT and LAAT of passenger subtrains. Let us suppose set $W$ has $L$ elements and defined as $W = \{w_1, w_2, \ldots, w_L\}$. For a given physical train, EADT of its first subtrain departing from its origin station (i.e., the true origin station of the physical train) toward the first intermediate station ($EADT^{w_1}$) is set equal to the beginning of the time horizon $T$. EADT of the second subtrain of the same physical train traveling from the first intermediate station to the second intermediate station ($EADT^{w_2}$) equals the beginning of time horizon $T$ plus unimpeded travel time from the true origin station to the first intermediate station. The procedure continues until the last subtrain of the same physical train and repeats across all physical trains. LAAT is computed in a similar but backward manner: we set LAAT of the last subtrain of a physical train traveling from the last intermediate station to the train’s final destination station ($EADT^{w_L}$) equal to the end of the planning time horizon $T$, and then compute LAAT backwards iteratively for the second-to-last subtrain, third-to-last subtrain, etc., until the first subtrain of the same physical train, using the unimpeded travel time between station pairs. We again repeat the procedure for all physical trains.

Passengers departing from the origin of station pair $w$ towards the destination of station pair $w$ consist of those with the destination of station pair $w$ as their true destination station, as well as those whose final destination station is in the same direction but farther than the destination of station pair $w$. For the sake of simplicity, these passengers are titled "passengers departing from the origin of station pair $w"", which is applied to the remaining of the paper. It is therefore clear that the calculation of passenger schedule delay at a given station involves passengers leaving this station for multiple destination stations in the same direction.

Among passengers departing from the origin of station pair $w$, those who desire to leave prior to the departure of subtrain $r_1^w$ will take this first subtrain. Some of the passengers who prefer to depart between departure times of subtrains $r_1^w$ and $r_2^w$ will also take the first subtrain. Other passengers whose PDT is in this interval will take the second subtrain. This continues until the end of the day. Passengers whose PDT is later than the departure of $r_L^w$, i.e., the last subtrain, will take the last subtrain. This process is justified by the fact that doing so gives minimum schedule delay cost for each passenger. Other things (e.g., fare) being equal (as we assume in this paper), minimizing schedule delay is equivalent to minimizing generalized travel cost for
each traveler. The different passenger schedule delay components for passengers departing from the origin of station pair \( w \) are formulated as (1.1)-(1.10). Below we describe the calculation of passenger schedule delay costs in great detail.

Because the effect of schedule inconvenience may depend on whether scheduled displacement is made towards an earlier or later train, we introduce two distinct cost coefficients, \( c_d^W \) and \( c_d^R \), which denote the respective unit cost ($/min) when an earlier and a later displacement is made by a traveler. In (1.1), we calculate total schedule delay cost of passengers who prefer to depart before \( u \), the departure time of the first subtrain \( r^W_1 \). These passengers will take subtrain \( r^w_1 \). We multiply the number of such passengers at the \( m \)th time interval \( (q^w_m) \) by the time difference between that time interval and subtrain \( r^w_1 \)'s departure time, \( u - \frac{m-1+m}{2} \), where we use the midpoint of the time interval \( (\frac{m-1+m}{2}) \) as the reference time point. We then sum over all time intervals from the beginning of the day \( (m=1) \) until \( u \) \((m=u)\), which gives the total schedule delay of passengers whose PDT is earlier than \( u \). Because these travelers incur schedule displacement towards a later train, total schedule delay is multiplied by \( c_d^R \).

In (1.2), we calculate schedule delay cost of passengers whose PDT is between \( u \) and \( u' \), where \( u' \) is the departure time of the second subtrain, \( r^W_2 \), and take the first subtrain \( r^w_1 \). As mentioned above, the traveler self-selection process of which train to take is based on minimize the schedule delay cost. For passengers whose PDT is between \( u \) and \( u' \), there exists a critical point where travelers have the same schedule delay costs whether taking the first or the second subtrain, and therefore would be indifferent to taking either train. Let that point be \( t_c \). Then the indifference implies \( c_d^W \ast (t_c - u) = (u' - t_c) \ast c_d^R \), which yields \( t_c = \frac{c_d^W u + c_d^R u'}{c_d^W + c_d^R} \).

Similar to (1.1), we multiply the number of passengers at each appropriate time interval \( m \) by the corresponding schedule shift \( \frac{m-1+m}{2} - u \) and cost coefficient \( c_d^W \) (displacement towards an earlier train), and sum over all \( m \) from \((u + 1)\) to \( u + 1 \) \( (\frac{1}{c_d^W + c_d^R}u') \) (using the floor function as an approximation). The resulting term, \( \frac{C_{u,R}^W}{u} \), gives the total schedule delay cost of passengers whose PDT is between \( u \) and \( u' \) and take \( r^W_1 \) given the departure time of the second subtrain \( r^W_2 \).

In (1.3), we relate the passenger schedule delay cost calculated in (1.2) to the starting arc of the second subtrain, \( x^w_{o^w,j,u',u'} \) and sum over all starting arcs of the second subtrain which do not violate the subtrain orders \((i.e., u < u')\). As a result, (1.3) represents the total schedule delay cost of passengers who prefer to depart after \( u \) and end up boarding subtrain \( r^W_2 \). The same logic is employed in (1.4)-(1.10) to derive mathematical representations of schedule delay costs for passengers boarding the intermediate and last subtrains.

Schedule delay cost calculation for the first subtrain

\[
C_{u,L}^{r^W} = \sum_{m=1}^{u} c_d^W q_m^w \left( u - \frac{m - 1 + m}{2} \right)
\tag{1.1}
\]

\[\forall w \in W, \forall u \in \{u | (o^w,j,u,v) \in \Psi^{p, r^W_1}\}\]

\[
C_{u,R}^{r^W} = \sum_{m=u+1}^{w} c_d^R q_m \left( \frac{m - 1 + m}{2} - u \right)
\tag{1.2}
\]

\[\forall w \in W, \forall (u,u') \in \{(u,u') | (u < u'), (o^w,j,u,v) \in \Psi^{p, r^W_2}, (o^w,j,u',v') \in \Psi^{p, r^W_1}\}\]

\[
C_{u,R}^{r^W} = \sum_{u | u < u'} \frac{C_{u,R}^{r^W}}{x^w_{o^w,j,u',u'}} \tag{1.3}
\]

\[\forall w \in W, \forall u \in \{u | (o^w,j,u,v) \in \Psi^{p, r^W_2}\}\]

Schedule delay cost calculation for an intermediate subtrain

\(1.4\)
\[ C_{u,k}^w = \sum_{m=1}^{u} \frac{c_d^R d_m}{c_d^R + c_d^m} (u - \frac{m-1+m}{2}) \]
\[ \forall w \in W, \forall n = 2, 3, \ldots, N - 1, \forall (u, u') \in ((u, u')| u' < u, (o^w, j, u', v') \in \Psi^{p, r_w}_{n-1}, (o^w, j, u, v) \in \Psi^{p, r_w}_n \} \]
\[ C_{u,k}^w = \sum_{u'|u < u'} C_{u,k}^w \times x_{o^w,j,u',v'}^{w-1} \]
\[ \forall w \in W, \forall n = 2, 3, \ldots, N - 1, \forall (o^w, j, u, v) \in \Psi^{p, r_w}_n \} \]

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

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

\[ \sum_{m=1}^{u} \frac{c_d^R d_m}{c_d^R + c_d^m} (u - \frac{m-1+m}{2}) \]
\[ \forall w \in W, \forall n = 2, 3, \ldots, N - 1, \forall (u, u') \in ((u, u')| u' < u, (o^w, j, u', v') \in \Psi^{p, r_w}_{n-1}, (o^w, j, u, v) \in \Psi^{p, r_w}_n \} \]

The objective function on the passenger side is then formulated as (2.1), where we relate the total schedule delay costs of passengers boarding each subtrain to the starting arcs of the subtrain, and then sum over all subtrains (the first three summation terms in (2.1)). Note that this is a quadratic objective function as \( x \) variables are multiplied by \( C \) variables, each of which entails a number of \( x \) variables. Recall that each passenger subtrain which arrives at a station is assigned to use one unit of time for sinking into the artificial sink node \( e^r \).

The last term in (2.1) penalizes subtrain \( r \) for staying more than \( (t_{r_{min}} + 1) \) in the intermediate station, where \( t_{r_{min}} \) is the minimum layover time of subtrain \( r \) at a station. If a fixed amount of waiting is required at each intermediate station, one can set both \( t_{r_{min}} \) and \( t_{r_{max}} \) to be that amount, and the preprocessing stage only generates \( y \) variables which meet the fixed layover time. In doing so, the fourth component in (2.1) can be dropped as the total layover time at stations becomes a constant.
\[
\text{Min} \quad \sum_{(o^w,f,j,u,v) \in \Psi \cap \Psi} (C_{u,l}^w + C_{u,R}^w) x_{o^w,f,j,u,v}^w + \sum_{n=2,3,\ldots,N-1} \left( C_{u,l}^w + C_{u,R}^w \right) x_{o^w,f,j,u,v}^w \\
+ \sum_{(o^w,f,j,u,v) \in \Psi \cap \Psi} (C_{u,l}^w + C_{u,R}^w) x_{o^w,f,j,u,v}^w + \sum_{(r,r) \in \Z^P} d^r \left( \hat{t} - (i^r_{\text{min}} + 1) \right) y_{t,\hat{t}}^{r,p}
\]

(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

\[
\sum_{(o^w,f,j,u,v) \in \Psi \cap \Psi} x_{o^w,f,j,u,v}^w = 1 \quad \forall n = 1,2,\ldots,N, \forall w \in W 
\]

(2.2)

\[
\sum_{(d^w,e^w,u,v) \in \Psi \cap \Psi} x_{d^w,e^w,u,v}^w = 1 \quad \forall n = 1,2,\ldots,N, \forall w \in W 
\]

(2.3)

\[
\sum_{(a,l,t,u) \in \Psi \cap \Psi} \sum_{(i,j,t,v) \in \Psi \cap \Psi} x_{a,l,t,u}^i \quad \forall r \in R^p, i \in \{B|i \neq o^w\}, t \in T 
\]

(2.4)

\[
\sum_{(d^w,e^w,u,t) \in \Psi \cap \Psi} x_{d^w,e^w,u,t}^r \quad \forall r, \hat{r} \in R^p, (r,\hat{r}) \in Z^p, t \in T 
\]

(2.5)

\[
\sum_{(o^w,f,j,t) \in \Psi \cap \Psi} \sum_{(t,\hat{t}) \in \Psi \cap \Psi} y_{t,\hat{t}}^{f,p} \quad \forall r, \hat{r} \in R^p, (r,\hat{r}) \in Z^p, t \in T 
\]

(2.6)

\[
x_{i,j,u,v}^f y_{t,\hat{t}}^{r,p} \in \{0,1\}
\]

(2.7)

side constraints

\[
\sum_{r \in R^p} x_{i,j,u,v}^r \leq b_i^r \quad \forall i \in B, t \in T 
\]

(3.1)

\[
\sum_{(i,j,u,v) \in \Psi \cap \Psi} x_{i,j,u,v}^f + \sum_{r \in R^p} x_{i,j,u,v}^r \leq v_i^f \quad \forall (a,t) \in T 
\]

(3.2)

\[
\sum_{r \in R^p} x_{a,j,u,v}^r \leq b_i^r \quad \forall i \in B, t \in T 
\]

(3.3)

In the linear network constraints, (2.2) and (2.3) guarantee unique departure and arrival for each subtrain. Flow conservation constraint (2.4) ensures continuity 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 may 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.

\[\text{Note that it is sufficient to write constraints (2.2)-(2.3) for the first subtrain of each physical train as flow conservation constraints will guarantee unique departure of other subtrains. However, establishing this constraint for all subtrains generates extra cuts, thereby lowering running time.}\]
Here, constraint (2.3) is not essential. However, we still keep the constraint because it can generate extra cuts while solving the integer program, therefore helping reduce the problem solving 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 all potential transitions of trains (in both directions) between blocks, which could be made via each cell within each transition window \([t + 1 - \delta, t + 1 + \delta]\), the length of which depends on the lagging and leading transition parameters \( \delta \) and \( \varepsilon \). It controls the transitions associated with each time window by setting the sum of all involved arcs equal to or less than the capacity of the cell at the associated time period. Constraints (3.3) and (3.4) manage headway, i.e., the minimum physical separation distance between a pair of leading and following subtrains measured in track blocks, in each direction. Overall, (2.2)-(3.1) are essential constraints in any train scheduling model and commonly used in the literature. On the other hand, constraints (3.2)-(3.4) are borrowed from Harrod (2011).

In solving for the passenger train schedules, the order of subtrains must be preserved, i.e. later departure of subtrain \( r_{n-1}^w \) than of subtrain \( r_n^w \), \( n = 2, \ldots, N \) is not allowed. To this end, we penalize any combination of the starting arcs of two consecutive subtrains which violates the order of subtrains (i.e. \( u' \geq u \)) by a large number \( M \), where \( u' \) and \( u \) are departure times of subtrains \( r_{n-1}^w \) and \( r_n^w \), respectively. Maintaining the order among subtrains essentially ensures maintaining the order among physical trains. The associated term, as shown in (4), is added to the objective function (2.1). We use the maximum possible passenger schedule delay cost for the value of \( M \). This is done by multiplying the number of passengers in each time period by the maximum possible schedule shift for those passengers, which is equal to the larger time distance between the time period under study and the two ends of the planning horizon, and then sum over all time periods. This value could be also considered as an upper-bound of (2.1). When (4) is in place, there is no need to set ad-hoc values for earliest allowed departure and latest allowed arrival times for each physical train, and the mathematical program will take care of the train order and find the schedule that minimizes total passenger schedule delay.

\[
\sum_{n=2,3,\ldots,N} \sum_{(u,u')\mid u'\leq u} M \times x_{o^w,j,u',u'}^r \times x_{o^w,j,u,u'}^r
\]

### 3.2.2 Lower level: freight train scheduling

The lower level problem is to develop freight train schedules, conditional on the passenger train timetable produced from the upper level. Different from the passenger side, here we do not need to decompose each physical freight train into subtrains, as freight trains do not have scheduled layover at intermediate stations. We assume that each freight train \( r \in R^f \) travels from the origin node of station pair \( w \) (i.e., \( o^w \)) to the destination node of station pair \( w \) (i.e., \( d^w \)), where \( w \) could be any arbitrary pair of stations. Stopping is possible when incurred by line capacity constraints. One can also specify different speeds for various freight trains, which is considered in setting up the feasible arc set of each train \( \Psi^f/\). Since solving the lower level freight side problem is subject to passenger train schedules from the upper level, careful attentions need to be paid to supplying input parameters. For each passenger subtrain \( r \in R^p \), we set the earliest departure time \( EDAT^r \) and the latest arrival time \( LAAT^r \) equal to the departure time from origin and the arrival time at the destination found at the first level. In this way, the preprocessing stage at the lower level will generate the set of feasible path arcs for passenger subtrains \( \Psi^p/\) identical to the set of arcs characterizing the optimal passenger subtrain schedules. Therefore this imposes passenger train schedules as constraints to the freight train scheduling problem.

In the US practice, a freight train is dispatched whenever the train receives enough load (Cordeau et al., 1998). Given that a freight train’s earliest allowed departure time, or its estimate, is known, the central planner would want to dispatch the train as early as possible. In general, dispatching a freight train at an earlier time increases the probability of on-time arrival at the destination, and makes available more capacity to subsequent freight trains. In addition to this, it is of interest to the freight side to incur less en-route delays as they not only lead to late arrival at the destination but also impose further operating costs, including fuel cost, driver cost, etc., to the freight railroad. While the effect of excess demand in a roadway system is manifested in congestion delay to the travelers, additional freight demand in a railway system operating near capacity results in the inability of the freight operator to obtain the slots that it needs (Nash and Sansom, 1999).
opportunity cost, more specifically loss of operating revenue due to demand that is foregone to the freight operator. Following the discussion, the full cost minimization problem on the freight side is modeled as follows:

\[
\begin{align*}
\text{Min} & \quad \sum_{r \in R'} \left( c^f_e (u - EADT^r) - c^p_i \right) x^r_{o^w, j, u, v} + \sum_{r \in R'} c^r f x^r_{i, j, u, v} \\
& \quad + \sum_{(i, j, u, v) \in \Psi f, r | i = j} c^m (v - t^r) x^r_{i, j, u, v} \\
\text{s.t.} & \quad \sum_{(o^w, j, u, v) \in \Psi f, r} x^r_{o^w, j, u, v} \leq 1 \quad \forall \{r \in R'\} \\
& \quad \sum_{(d^w, e^r, u, v) \in \Psi f, r} x^r_{d^w, e^r, u, v} \leq 1 \quad \forall \{r \in R'\} \\
& \quad \sum_{(a, i, u, v) \in \Psi f, r} x^r_{a, i, u, v} = \sum_{(i, j, u, v) \in \Psi f, r} x^r_{i, j, u, v} \quad \forall r \in R', i \in \{B | i \neq o^w\}, t \in T \\
& \quad \sum_{r \in R'} \sum_{(d^w, e^r, u, v) \in \Psi f, r} x^r_{d^w, e^r, u, v} = \sum_{(o^w, j, u, v) \in \Psi f, r} x^r_{o^w, j, u, v} \\
& \quad (2.2)-(2.7) \quad \forall r \in R^p, \text{given } EADT^r \text{ and } LAAT^r \text{ found from the upper level} \\
& \quad (3.1)-(3.4) \quad \forall r \in R_s, \text{given } EADT^r \text{ and } LAAT^r \text{ found from the upper level} \\
& \quad x^r_{i, j, u, v} \in \{0, 1\} \\
\end{align*}
\]

The first term in (5.1), \( c^f_e (u - EADT^r) \), represents the freight train delay cost due to late departure. Given a prescribed maximum number of freight trains that desire to be scheduled (i.e., total freight train demand), foregone demand cost is represented by the negative of the number of freight trains that can be actually scheduled, times the unit foregone demand cost per freight train, \( c^p_i \), the greater the number of freight trains that can be actually accommodated, the smaller the foregone demand cost. The second term in (5.1) captures en-route delay costs associated with waiting arcs \((i = j)\). The third term considers the freight train operating cost and the associated freight time value while the freight trains are moving.

The freight side cost minimization is also subject to linear network and side constraints. Constraints (5.2) and (5.3) enforce unique departure and unique arrival for each freight train, respectively. In contrast to (2.2) and (2.3), inequality is possible in (5.2) and (5.3) because 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 ending arcs in the opposite direction. As a result, equal number of freight trains run in both directions. Constraint (5.6) imposes passenger train schedules found from the upper level. 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 passenger train schedules are now given. Constraint (5.8) regulates that the \( x \) variables can only take value 0 or 1.

Under the constant speed assumption, the third term in (5.1) is constant and can be removed from the objective function. As a consequence, we only need to minimize the objective function (6):

\[
\begin{align*}
\text{Min} & \quad \sum_{r \in R'} \left( c^f_e (u - EADT^r) - c^p_i \right) x^r_{o^w, j, u, v} + \sum_{r \in R'} c^r f x^r_{i, j, u, v} \\
\text{s.t.} & \quad (5.2)-(5.8) \\
\end{align*}
\]

4 Solution approach

This section discusses different model reformulations to solve the upper level (passenger side) problem of the two-level optimization model, in order to improve the computational performance. For the lower level
(freight side), the linear integer program can be conveniently solved using off-the-shelf techniques. As is made clear in sub-Section 3.2.1, the upper level's objective function involves the products of \( x \) and \( C \). Note that each \( C \) expression in (1.3), (1.5), (1.7), and (1.9) involves \( x \) variables, each of which denote the departure time of a neighboring subtrain. Therefore, the passenger side scheduling is a Quadratic Integer Programming (QIP) problem, which can be abstracted as (7.1)-(7.4). Because each \( x \) variable appearing in the objective function represents a starting arc of a subtrain, each quadratic term in the objective function represents a combination of stating arcs of two consecutive subtrains. QIPs are in general NP-hard, and remain so even for relaxed QIP problems with continuous variables, where matrix \( H \) in (7.1) is indefinite (Pardalos and Vavasis, 1991; Sahni, 1974). Our numeric analyses reveal that matrix \( H \) always have both positive and negative eigenvalues. Hence, the computational complexity order of solving the passenger train scheduling problem (7.1)-(7.4) falls into the NP-hard category. Considering reasonable length of the planning horizon, and the number of passenger trains and line blocks can lead to a large number of decision variables and constraints, thus requiring significant computation time to solve the QIP problem. The purpose of this section is to explore ways to reduce the computational efficiency.

\[
\text{Max } f(x) + \frac{1}{2} x^T H x \quad (7.1)
\]

\[
s.t.
A_{ineq} x \leq b_{ineq} \quad (7.2)
\]

\[
A_{eq} x = b_{eq} \quad (7.3)
\]

\[
x \in \{0,1\} \quad (7.4)
\]

The first remedy is dropping the term involving big \( M \), (4), from the objective function and instead introducing a new set of constraints (8) to ensure subtrain orders. This is because using big \( M \) creates large difference in value, or even in magnitude, among the different terms in the objective function, and 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). For each pair of consecutive subtrains (servicing the same station pair \( w \), constraint (8) is imposed on the departure of the second subtrain. The starting arc of subtrain \( r_n^w \) plus the sum of starting arcs of the previous subtrain (i.e., \( r_{n-1}^w \)) departing later than subtrain \( r_n^w \), must be equal to or less than one. In this way, only \( x \) variables that are associated with earlier departure of the first subtrain than the second subtrain will be assigned value 1, therefore ensuring the order among subtrains. In addition, the new constraints (8) introduce new cuts which also help improve computational efficiency.

\[
\sum_{(o,w,j,u',v') \in \psi^{p,r_{n-1}^w}|u'L} x_{o,w,j,u',v'}^{r_n^w} + x_{o,w,j,u,v}^{r_n^w} \leq 1 \quad (8)
\]

\[
\forall w \in W, \forall n = 2,3, \ldots, N, \forall u|(o^w, j, u, v) \in \Psi^{p,r_n^w}.
\]

A second remedy is to reformulate the problem (7.1)-(7.4) to an equivalent binary linear program using transformation (9.1)-(9.4). In the reformulated problem, we substitute each combination of starting arcs of two consecutive subtrains in (2.1), which does not violate the order of trains, with a new 0-1 binary variable, \( z_{u',u}^{r_n^w} \), as is in (9.1). For each new variable, three inequality constraints (9.2)-(9.4) need to be added to maintain the relationship among the variables. Constraints (9.2) and (9.3) express that setting the value of at least one of the \( x \) variables to zero obliges the associated \( z \) variable to be equal to zero. Constraint (9.4) ensures that the value of the \( z \) variable equals one if and only if both associated \( x \) variables are equal to one. Because of the reformulation, the size of the problem becomes significantly larger, in terms of both the numbers of variables and constraints. However, the problem is no longer quadratic but linear, which helps improve the computational efficiency in solving the original program.

\[
z_{u',u}^{r_n^w} = x_{o,w,j,u',v'}^{r_n^w} x_{o,w,j,u,v}^{r_n^w} \quad (9.1)
\]

\[
\forall w \in W, \forall n = 2,3, \ldots, N, \forall (u', u)|u' < u, (o^w, j, u', v') \in \Psi^{p,r_{n-1}^w}, (o^w, j, u, v) \in \Psi^{p,r_n^w}.
\]
Although reformulation (9.1)-(9.4) suggests improved computational performance due to linearization, our computational experiments show that large problems can still not be solved within a reasonable amount of time. This may be attributable to the introduction of many more new binary variables and constraints. In particular, constraint (9.4) significantly affects the solving time and omitting this set of constraints considerably lowers the computational complexity and problem solving time. One could use the standard Lagrangian relaxation approach, which progressively improves the obtained lower bound of the objective function through the subgradient algorithm until reaching an acceptable gap between the lower bound and upper bound values. In our case, the upper bound can be quickly found by setting the value of \( z \) variables associated with starting arcs of each pair of consecutive subtrains equal to one. However, the Lagrangian relaxation remains an unsatisfactory strategy for our problem because of the very slow rate in lower-bound improvement. Basically, the subgradient value of any violated constraint equals at most one, which is quite insignificant in contrast to the objective value. Consequently, the convergence of the subgradient algorithm will be largely affected by Lagrangian multipliers’ initial values. In this study, we instead propose to introduce new constraints (10) in place of (9.4), which take advantage of unique problem structure. As passenger subtrains are labeled “must run”, one of the starting arcs for each subtrain must equal one. Therefore, one and only one of the \( z \) variables associated with each pair of consecutive subtrains must equal one, as shown in (10).

Compared to (9.4) which is a standard linearization technique not considering the information of the problem structure, (10) exploits such information and subsequently represents the same characteristics with much fewer constraints.

\[
\sum_{(u',u) | u' < u, (o^w, j, u', v') \in \Psi^w, n = 1} z^w_{n-1} f^w_{n-1} = 1
\]

(10)

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

Summing up, four model formulations are considered for the passenger side train scheduling problem:

**F1:** \( \text{Min } (2.1) + (4) \)
\[ \text{S.t. } (2.2)-(2.7) \]
\[ (3.1)-(3.4) \]

(11)

**F2:** \( \text{Min } (2.1) \)
\[ \text{S.t. } (2.2)-(2.7) \]
\[ (3.1)-(3.4) \]
\[ (8) \]

(12)

**F3:** \( \text{Min } (2.1)' \)
\[ \text{S.t. } (2.2)-(2.7) \]
\[ (3.1)-(3.4) \]
\[ (8) \]
\[ (9.2)-(9.4) \]
\[ z \in [0,1] \]

(13)

**F4:** \( \text{Min } (2.1)' \)
\[ \text{S.t. } (2.2)-(2.7) \]
\[ (3.1)-(3.4) \]

(14)
where \( (2.1)' \) denotes a reformulation of \( (2.1) \) by substituting the product of \( x \)'s by \( z \)'s, as is in \( (9.1) \). As already mentioned, \( F_2 \) is expected to solve the passenger side train scheduling problem faster than \( F_1 \), as \( F_2 \) takes the advantage of new cuts and smaller magnitude differences for the different terms in the objective function. By transforming the QIP problem to linear IP problems, standard techniques for solving linear IP problems can be employed, which leads to even more efficient computation (Chaovalitwongse et al., 2004; Klotz and Newman, 2013). Therefore, \( F_3 \) and \( F_4 \) are likely to show improved computational performance compared to \( F_1 \) and \( F_2 \), despite more variables and constraints introduced in \( (9.2)-(9.4) \). We further expect \( F_4 \) to outperform \( F_3 \), as the number of constraints in \( (10) \) is much lower than in \( (9.4) \). The 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 Numerical analysis

This section demonstrates implementing the two-level train scheduling model on shared use rail corridors. We start with determining model input parameter values. Then detailed presentation of solving a sample problem is provided, including investigation of the impact on model results of train speed heterogeneity and freight train delay tolerance levels. The last part of our numerical analysis is a case study of the prospective Chicago-St Louis higher speed rail corridor, where freight and 110-mph passenger trains will be running on a shared single track (with sidings) which is currently being upgraded, with infrastructure improvement expected to complete by 2017.

5.1 Modeling parameters

The first step in implementing the model is to determine the model input parameter values. On the passenger side, the model requires a temporal distribution of passengers’ Preferred Departure Time (PDT), and the time value of passenger schedule delay \( (c^L_d \text{ and } c^R_d) \). To our knowledge, the existing literature on passenger PDT is limited, probably because passenger PDT can only be obtained through specific surveys. Because obtaining passenger PDT profiles is beyond the scope of our study, in this research we use an existing passenger PDT profile estimated by Cascetta and Coppola (2012) for Italian intercity travel, as shown in Figure 6, and assume that the distributions of passenger PDTs for all OD pairs are identical. For value of passenger schedule delay, we follow Corman and D’Ariano (2012), who specifically estimate values of passenger travel time for rail transportation, and use 1.5 times of passenger line-haul travel time cost for schedule delay cost. We use the US Department of Transportation’s reference passenger value of line-haul travel time for rail, which equates $32.9/hr and $58.9/hr for personal and business trips respectively (US DOT, 2011). Following IDOT (2012), we assume that 7% of total trips are for business purposes in the subsequent analysis. The average passenger values of in-vehicle travel time and schedule delay time \( (c^L_d = c^R_d) \) are therefore $34.7/hr and $52/hr, respectively.
The freight side model uses three unit cost parameters: foregone demand cost ($c_f^i$), departure delay cost ($c_d^i$), and en-route delay cost ($c_e^i$). Departure delay cost $c_d^i$ is considered equal to freight value of time, as late departure does not impose significant operating costs to a train. In contrast, en-route delay cost $c_e^i$ encompasses train operating cost while stopping and freight value of time (in other words, $c_e^i$ is part of $c_f^i$).

The estimation of freight foregone demand cost requires revealing the value placed on the slot by an alternative user (Johnson and Nash, 2005). Johnson and Nash (2005) estimate the opportunity cost of a slot for freight trains in the UK equal to £0.0149/ton-mile. According to the Association of American Railroads (AAR, 2012), in 2010 Class I Railroads hauled 1,691 billion ton-miles of freight, with a total revenue of $58.41 billion. The average operating revenue was $0.035/ton-mile, which is in the same range of that from Johnson and Nash (2005). Multiplying $0.035/ton-mile by the average tonnage hauled by each train (3585 tons, based on AAR (2012)), the estimated freight foregone demand cost is $123.8/mile. Foregone demand cost per freight train $c_f^i$ can then be conveniently calculated by multiplying the length of the corridor by $123.8$/mile.

Freight Value of Time (VOT) is an important issue in rail transportation, but has been largely neglected in train schedule modeling literature. To our knowledge, the latest rail-specific freight VOT estimation in the United States is conducted by Vieira (1992) (Feo-Valero et al., 2011; Zamparini and Reggiani, 2007). Using Vieira’s estimation and an average of 3% inflation rate per year, we obtain rail freight VOT in 2010 to be $0.955$/ton-hr. Multiplying that value by the average tonnage hauled by a typical Class I railroad train (3585 tons/train), $c_f^i$ is estimated to be $3424.1$/train-hr.

Our estimate of freight train operating costs is also based on AAR (2012), by dividing the reported total operating expenses by the total train-hours over all Class I railroads. We discount the estimated operating cost by 10%, given that fuel consumed while waiting en route burn much less fuel (indeed close to zero), and the fact that fuel accounts for about 8% in total operating cost (Mizutani, 2004). This leads to an estimated en-route operating cost while stopping at $1633.1$/train-hr. Further adding freight VOT $c_f^i$, $c_e^i$ is estimated to be $5057.2$/train-hr.

5.2 A Sample problem

The sample problem considers a single-track rail line with 11 blocks and two OD pairs, one for each direction. As there is no intermediate station along the rail line, each physical passenger train is equivalent to a single subtrain in the formulation. Six of the 11 blocks are equally distanced single-track segments and the
other five, inserted between the single-track segments, are equally distanced, shorter double-track segments, i.e. segments each with a siding. Each single-track segment is 18 miles long (except the first and the last segments which are 19 miles long); whereas, the length of each siding is 2 miles. Because the total length of the line is 120 miles (<400 km), we assume that passenger PDT distribution in each direction follows the left-hand-side profile in Figure 6.

We solve the train scheduling problem with different even numbers of passenger trains (to keep the same number of trains in both directions) dispatched between 5:00 AM and 9:30 PM (thus in total 16.5 hours). For the most frequent passenger rail service case, i.e., 6 trains running each way, we assume a daily demand of 760 and 710 passengers traveling in the two directions. As passenger demand is known to be sensitive to train frequency, we follow Adler et al. (2010) and use a frequency elasticity of 0.4 to come with estimates of daily passenger demand with lower passenger train frequencies. Once the total number of passengers is determined, we proportionately adjust the passenger PDT profile between 5:00 AM and 9:30 PM. The operating speed of passenger trains is set at 120 mph.

We assume that freight trains run at 60 mph, which is faster than the current average speed of freight trains in the US. This is intended to capture the potential speed increase of freight services after the rail line is upgraded to accommodate 120 mph HSR service. While passenger trains can be freely scheduled throughout the day (as long as the order of trains is maintained), each freight train has a predetermined earliest allowed departure time. However, information about the earliest allowed train departure time may not be available at the strategic planning stage. In this study (the sample problem as well as the larger problem later), we assume a uniform distribution of freight trains' earliest allowed departure time, with one-hour interval, throughout the time horizon. We follow Sahin et al. (2008) and assume that the maximum delay tolerance for a freight train is 50% of the train's unimpeded (i.e. unstopped) travel time along the line. Therefore, the latest allowed arrival time of a freight train is its earliest allowed departure plus 1.5 times its unimpeded travel time. Certainly, this number can be changed to reflect different tolerance levels, which we investigate later in sub-Section 5.4. The earliest departure time of the first freight train in a day is the beginning of the 16.5-hour planning horizon. Without any further constraints a maximum of 15 freight trains can be scheduled in each direction (noting that unimpeded running time from one end to the other on the line is 2 hours). Same as in Harrod (2011), the minimum headway $h^r$ $(\forall r \in R)$ is set equal to one block.

The choice of time resolution unit plays a key role in determining the computation time. According to Sahin et al. (2008), considering train movements by 5-min interval can provide sufficient accuracy for train scheduling at the strategic level. We therefore choose five minutes as the length of a time resolution unit. On the other hand, passenger PDT profiles are usually identified at coarser levels in practice, e.g., in our case every 15 minutes. To reconcile these two time resolutions, we make the assumption that passengers are uniformly distributed within each 15-minute period. Thus the numbers of passengers in each 5-min interval is calculated by dividing the numbers of passengers in the corresponding 15-min period by three.

The numerical experiments from solving the passenger side problem confirm our expectations of the computational performance under different model formulations discussed in Section 4. The big $M$ approach (F1) shows very limited capability, failing to solve the passenger train scheduling problem due to memory overflow. In the scenario with two passenger trains in each direction, CPLEX produces the optimal solution under F2 in 46.25 hours. Solving time is significantly improved when the quadratic IP problem is linearized (F3 and F4). In particular, F4 can produce optimal results for up to six passenger trains scheduled in each direction within one minute, as shown in Table 3. Given its superior performance, in what ensues we always use F4 to solve the passenger side trains scheduling problem. On the freight side, optimal solution can always be found in very short periods, in the range of 1-10 seconds.

<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>26 sec</td>
</tr>
<tr>
<td>Sc. 3</td>
<td>4 passenger trains</td>
<td>38 sec</td>
</tr>
<tr>
<td>Sc. 4</td>
<td>5 passenger trains</td>
<td>39 sec</td>
</tr>
<tr>
<td>Sc. 5</td>
<td>6 passenger trains</td>
<td>52 sec</td>
</tr>
</tbody>
</table>

Table 3: Running time for solving the passenger side problem using F4
Table 4 reports the departure time of passenger trains along with passenger demand under different scenarios. We observe that, in all scenarios with more than one passenger train per day in each direction, there will be one train scheduled during the morning and afternoon peaks, respectively. Detailed train schedules in time-space diagrams are provided in Appendix. As we assume the same distributions for passenger PDT for both ODs, having symmetric train departure times in each direction is expected. The asymmetric departure times under the one-train scenario is caused by rounding passenger numbers. Symmetric train departures suggest that two trains departing at the same time and running in opposite directions will meet at the exact midpoint of the line (as shown in the figures in Appendix), which is a siding in our line configuration.

### Table 4: Comparison of departure times for simultaneous and sequential approaches

<table>
<thead>
<tr>
<th>Passenger trains in each direction</th>
<th>Passengers</th>
<th>NB</th>
<th>SB</th>
<th>Passenger train departure time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>381</td>
<td>407</td>
<td>(16:25,12:15)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>476</td>
<td>509</td>
<td>(8:10,8:10)(17:05,17:05)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>549</td>
<td>587</td>
<td>(8:10,8:10)(15:45,15:45)(18:10,18:10)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>663</td>
<td>709</td>
<td>(7:20,7:20)(9:20,9:20)(12:50,12:50)(16:00,16:00) (18:30,18:30)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>710</td>
<td>760</td>
<td>(7:15,7:15)(9:10,9:10)(12:20,12:20)(15:10,15:10) (17:00,17:00)(18:50,18:50)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 7 plots the total and per passenger schedule delay cost, as a function of the number of passenger trains scheduled. While passenger demand increases with train frequency, total passenger schedule delay cost follows a similar decreasing trend as schedule delay cost per passenger. This marginal effect of schedule delay cost reduction diminishes with the number of passenger trains scheduled, with the most significant reduction occurring when increasing trains from one to two in each direction. In the least cost scenario (6 daily passenger trains in each direction), total passenger schedule delay cost in a day is still above $44000, with an average passenger bearing about $30, which is comparable to the passenger in-vehicle travel time cost ($34.7/hr * 1 hr = $34.7).

In Table 4, the total number of passengers for the one-passenger-train scenario is calculated based on arc elasticity of demand with respect to frequency: \( \text{Pax}_1 = \text{Pax}_2 (1 - \text{elasticity} \times 1/2) \), where \( \text{Pax}_2 \) denotes the total passenger demand with \( X \) -passenger \( -\text{train[s]} \). The distribution of passenger PDT in the one passenger train scenario is obtained by multiplying the PDT distribution of the two-passenger-train scenario by \( (1 - \text{elasticity} \times 1/2) \). In principle, the total number of passengers obtained from summing over all time periods should equal \( \text{Pax}_1 \). However, when only one passenger train is scheduled, the difference between the two summations is significant due to rounding errors. To make the two summations consistent, we consider a number of random time intervals, for each of which one passenger is added to the current number of passengers. This results in asymmetric distribution of PDT in the opposite directions.
Figure 7: Total and per passenger schedule delay cost

Table 5 documents the passenger schedule delay cost for each train and the numbers of passengers boarding each train with 1-6 passenger trains scheduled in each direction, as well as incremental passenger benefits when one more train is added to a given train service frequency. Due to the uneven distribution of passenger PDT in a day, each train bears different schedule delay costs. Not surprisingly, passenger trains whose departure times are close to morning and evening peaks carry more passengers than the other trains, therefore having higher 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 morning and evening peaks, are about as twice as train 5/6.

Table 5: 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>Trains in direction 1</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</td>
<td>4.65</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</td>
<td>4.17</td>
<td>4.46 (150)</td>
<td></td>
</tr>
<tr>
<td>Train 5</td>
<td>4.92 (145)</td>
<td>6.95</td>
<td>3.19 (74)</td>
<td>2.35 (63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 7</td>
<td>5.90</td>
<td>5.29</td>
<td>2.83 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 9</td>
<td>5.83</td>
<td>2.93 (124)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.87 (116)</td>
<td></td>
</tr>
<tr>
<td>Trains in direction 2</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</td>
<td>5.61</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</td>
<td>4.18</td>
<td>4.71 (158)</td>
<td></td>
</tr>
<tr>
<td>Train 6</td>
<td>5.67 (159)</td>
<td>6.85</td>
<td>3.37 (79)</td>
<td>2.45 (67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 8</td>
<td>6.63</td>
<td>5.48</td>
<td>2.93 (105)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 10</td>
<td></td>
<td>6.31</td>
<td>3.15 (132)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train 12</td>
<td></td>
<td></td>
<td>4.24 (125)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective function value ($000)</td>
<td>179.31</td>
<td>56.11</td>
<td>50.04</td>
<td>49.09</td>
<td>48.09</td>
<td>44.07</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>7.20</td>
<td></td>
</tr>
</tbody>
</table>

The last line in Table 5 provides an estimate of incremental passenger benefits, using the "rule of one-half", which is widely considered as a measure for user benefits in transportation appraisal (Nellthorp and Hyman,
Specifically, incremental passenger benefit by adding one more train is equal to \( \frac{1}{2} \Delta_{SD} (D_0 + D_1) \), where \( \Delta_{SD} \), \( D_0 \), and \( D_1 \) denote, respectively, the change in average schedule delay cost per passenger (for passengers using the train service) and total passenger demand without and with the additional train. This is performed for each station pair. Assuming a seating capacity of 378 for each train (IDOT, 2012), passengers can be fully accommodated under all scenarios except when there is only one passenger train scheduled in each direction.

Figure 8 shows the freight side costs, decomposed into costs associated with foregone demand, late departure, and en-route delay, plotted as a function of the number of passenger trains scheduled in each direction. As expected, the total cost increases as more passenger trains are scheduled, which demand a greater portion of the line capacity. When less than three passenger trains run in each direction, all freight trains will be able to operate on the line, although with departure and en-route delays. However, two freight trains will be forced out of service with the presence of 3-5 passenger trains, and this number increases to four when six passenger trains are scheduled in each direction. In the latter case, foregone demand becomes the most important cost component in the total. Whereas departure delay cost is relatively stable across all the six scenarios, en-route delay cost also has an increasing trend.

The results presented in the bottom of Table 5 and Figure 8 allow us to further compare the marginal passenger benefit gains, marginal freight cost increase, and the net cost effect while increasing the number of passenger trains (Figure 9). Net marginal benefit gains only occur when the number of passenger trains increases from one to two and from four to five (very slightly) in each direction. It is noteworthy that the relative importance of the marginal passenger and freight side costs is subject to the cost parameter values chosen, and could change if more accurate cost parameters become available.

![Figure 8: Shares of various components in freight-side total cost ($000)](image-url)
5.3 The impacts of speed heterogeneity

Speed heterogeneity is an important issue in shared use rail corridor operation performance (Dingler et al., 2009; Dingler, 2010; Sogin et al., 2013). Recall that our assumed HSR passenger train speed on the shared line is 120 mph. With the upgrade of rail tracks and related infrastructure in order to accommodate HSR services, it is also possible for freight trains to increase their operating speed.

In this sub-section, we suppose that passenger schedules are the same as in subsection 5.2. Recall that given the seating capacity suggested by IDOT (2012), all passengers will be accommodated except when only one passenger train is scheduled in each direction, in which case a small number of passengers will be unattended. Then, we investigate the cost impact of different freight train speeds, ranging from 12 mph to 120 mph. Due to the lack of data, we assume the same unit cost parameter values \( c_p, c_e, c_s \) under all speeds.

Figure 10 presents total freight side costs, as a function of the number of passenger trains scheduled. Each line corresponds to a specific freight train speed. Similar to the base case (60 mph), adding passenger trains raises total freight cost, with two exceptions: one is when freight trains run at 12 mph with two passenger trains in each direction; the other is when freight trains run at 120 mph with six passenger trains. These outliers may be interpreted by the fact that altering passenger train departure times results in better form of available line capacity to accommodate more freight trains or reduce freight train delays.

By and large, increasing freight train speed, which reduces speed difference between passenger and freight trains, leads to smaller freight side cost. This is consistent with the general finding in the literature (e.g. Dingler et al., 2009) that, regardless of train scheduling priority, the more heterogeneous the train traffic, the higher cost to freight trains. However, we observe two exceptions with five and six passenger trains. In these cases, freight trains running at 120 mph incurs higher total costs than at 75 mph. Scrutiny of the freight and passenger train paths reveals that, at 120 mph, the allowed operating time intervals for some freight trains have significant overlap with the passenger train timetable, leading to infeasible operation of these freight trains. On the other hand, this is less the case with freight trains running at 75 mph.

If a large number of passengers were unattended (i.e., passenger train seating capacity constraints are severe), then it may be more realistic to model the upper-level problem as minimizing the total cost of passenger schedule delay and unattended demand. This will generate a different schedule for passenger trains, and subsequently for freight trains. The impact on freight cost of speed heterogeneity as presented in this sub-section will also be different.

Departure delay cost is estimated based on freight time value, which is clearly independent of train speed. Foregone demand delay cost does not have an operational nature and hinges on the values that freight operators place on a slot. En-route delay cost consists of freight time value and train operating cost excluding fuel, which can also be reasonably approximated as constant.
The above argument is confirmed when we take a further look into the freight foregone demand under different freight train speeds, as shown in the left panel of Figure 11. The associated foregone demand cost is displayed on the right panel. When there are five or six passenger trains, some freight trains cannot be scheduled with 120 mph speed. In contrast, all freight trains can be accommodated if running at 75 mph. Apart from that, the number of freight trains shows a non-increasing trend with passenger train frequency across all speeds. In the extreme case that freight trains run at 12 mph, freight service will disappear entirely if scheduling four or more passenger trains in each direction. The foregone demand cost curves resembles those in Figure 10, reflecting a general increasing trend as passenger service increases.

Figure 11: Number of freight trains and foregone demand cost with different freight train speeds
Figure 12: Freight en-route delay cost with different freight train speeds

En-route and departure delay costs of freight trains are presented in Figure 12 and Figure 13, in terms of both total and per running train costs, with different speeds and passenger train services. When freight trains run at 120 mph, i.e., passenger and freight train speeds are homogeneous, no en-route delay will incur. However, no further general finding can be drawn from the total cost curves, as most of them are non-monotonic with respect to speed. En-route delay cost per running train seems to exhibit more consistent pattern: lowering the freight train speed elevates average en-route delay cost, if the two points with zero and two passenger trains on the 12 mph curve are excluded. In contrast, Figure 13 shows a less clear trend in both total and per freight train departure delay costs. Overall, foregone demand cost is the dominant component.

Figure 13: Freight departure delay cost with different freight train speeds
5.4 The impact of freight train delay tolerance levels

This sub-section examines the sensitivity of freight side cost to different train delay tolerance levels. Recall that the delay tolerance level specifies the guaranteed worst-scenario on-time performance on the arrival end. In the preceding sections, we use 0.5 as the multiplier of unimpeded trip time to bound the maximum departure and en-route delays (i.e., delay tolerance). In what follows we consider multipliers from 0.1 and 0.6 with 0.1 increments. Our numeric experiments (not shown here) suggest that 0.6 is an upper bound for freight train delay tolerance level, as setting tolerance levels beyond this threshold does not result in further improvement to the freight side cost.

Figure 14 illustrates freight foregone demand and the associated cost for various levels of delay tolerance. Given the passenger train service frequency, greater tolerance of delays leads to non-decreasing number of freight trains running on the line, therefore reducing foregone demand cost. This, however, needs to be weighed against the resulting degradation of freight train on-time performance, as shown in Figure 15, for a given level of passenger train service. On the other hand, increasing passenger train services generally reduces the number of running freight trains and increases foregone demand cost across all delay tolerance levels. If only one passenger train runs in each direction, no foregone demand will exist with tolerance level beyond this threshold does not result in further improvement to the freight side cost.

When six passenger trains are scheduled on the line, a delay tolerance level of 0.5 or 0.6 will incur $50000 foregone demand costs per day, which however is still one fifth of the cost with 0.1 delay tolerance level.

![Figure 14: Number of freight trains and corresponding foregone demand cost for various levels of freight train delay tolerance](image)

En-route and departure delay costs to freight trains under different delay tolerance levels are shown in Figure 15. Compared to the right panel of Figure 14 where the most strict delay tolerance level leads to the highest foregone demand cost, delay costs become larger as we progressively relax the delay tolerance level: when a 0.1 multiplier is used there is no train departure delay and en-route delay is also minimum; whereas the maximum delay costs are achieved when tolerance level is 0.6. The increase in late departure delay with respect to the delay tolerance level is most evident when four passenger trains are scheduled. Lowering the delay tolerance level leads to fewer freight trains running on the line. The resulting reduced congestion enables remaining freight trains to depart with lower departure delays. On the other hand, because delay related costs are much lower than the associated cost of foregone demand, lowering delay tolerance would still result in a net freight side cost increase. This is indeed consistent with the reality that freights railroads are willing to bear high delays but less willing to cut operations. Focusing on the sensitivity of delay cost to the number of passenger trains, the difference between the maximum and minimum en-route delay costs increases with more passenger trains. In contrast, the departure delay curves exhibit a less clear trend as we add more passenger trains.

En-route and departure delay costs to freight trains under different delay tolerance levels are shown in Figure 15. Compared to the right panel of Figure 14 where the most strict delay tolerance level leads to the highest foregone demand cost, delay costs become larger as we progressively relax the delay tolerance level: when a 0.1 multiplier is used there is no train departure delay and en-route delay is also minimum; whereas the maximum delay costs are achieved when tolerance level is 0.6. The increase in late departure delay with respect to the delay tolerance level is most evident when four passenger trains are scheduled. Lowering the delay tolerance level leads to fewer freight trains running on the line. The resulting reduced congestion enables remaining freight trains to depart with lower departure delays. On the other hand, because delay related costs are much lower than the associated cost of foregone demand, lowering delay tolerance would still result in a net freight side cost increase. This is indeed consistent with the reality that freights railroads are willing to bear high delays but less willing to cut operations. Focusing on the sensitivity of delay cost to the number of passenger trains, the difference between the maximum and minimum en-route delay costs increases with more passenger trains. In contrast, the departure delay curves exhibit a less clear trend as we add more passenger trains.
5.5 A larger problem

In this sub-section, we demonstrate the application of the model to a real-world size problem. We consider the Chicago-St Louis higher speed rail line, the track of which is owned by Union Pacific railroad and currently being upgraded to accommodate the 110 mph passenger rail service. The line schematic is shown in Figure 16, which is constructed based on the siding location and length information from Harnish (2013). The line is 285 miles long and consists of 14 single-track segments and 17 double-track segments. On the passenger side, while the prospective line has nine stations (including Chicago and St Louis), 90% of all OD trips will be among six stations, i.e., Chicago, Joliet, Normal, Springfield, Alton, and St. Louis, according to the Illinois DOT demand forecast (IDOT, 2012). In the subsequent analysis, only these six stations are considered as stops for the higher speed rail service. We further assume that passenger train layover time at intermediate stations is equal to one time period (i.e., 5 min).

We first consider a near term, 2015 scenario based on the IDOT aggregate passenger demand forecast. The aggregated daily passenger demand in 2015 is proportionately allocated to each OD pair using the current demand distribution among the OD pairs, to generate passenger trips for each OD, as shown in Table 6. Illinois DOT suggests that the higher speed rail corridor will be served by five passenger trains daily in each direction in the near term (IDOT, 2012). Similar as before, we assume that seating capacity of each train equals 378. Because six stations are considered on the corridor, each physical train is decomposed into five consecutive subtrains in the model formulation. Although the proposed maximum passenger train speed is 110 mph, we consider a discounted average speed of 90 mph due to train acceleration and deceleration.

We further assume that freight trains run at 30 mph. The IDOT 2012 study reveals that very few freight trains currently traverse the entire line end-to-end, and more freight traffic is near the south end (IDOT, 2012). Following the actual observed traffic, the following freight train demand is considered: 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 to St. Louis. The two directions have an equal number of freight trains for each OD pair. Similar to sub-Section 5.2, the earliest allowed departure times of freight trains

Figure 15: Freight en-route and departure delay costs under different delay tolerance levels

Figure 16: Schematic map of the higher speed rail line between Chicago and St. Louis
between each station pair are assumed uniformly distributed over the 16.5 hours planning time horizon, and the latest allowed arrival time for each freight train equals the earliest allowed departure time plus 1.5 times of unimpeded travel time. The other model parameters take the same value as those in Section 5.2.

<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>2082</td>
</tr>
</tbody>
</table>

We use the two-level strategic train scheduling model to solve for the optimal train schedules. In each direction, we sum all passengers originating from each station and distribute the calculated number over the course of the day using the right-hand-side graph of Figure 6. The top (passenger) level problem, which now includes 421627 variables, 52881 equality constraints, and 760498 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 estimated rail ticket price which is $39 between Chicago and St. Louis (IDOT, 2012), the schedule delay cost is substantial. On the freight side, only one train between St. Louis and Alton and one train between Springfield and Normal cannot be accommodated given the passenger train service. All train paths are illustrated in Figure 17. The breakdown of freight side cost is presented in the first bar in Figure 18.

Because passenger rail is projected to experience continuous demand growth in the future, as it has witnessed in the past decade, we also examine a mid-term, 2020 passenger demand scenario. OD trips in 2020 are interpolated using the IDOT 2015 and 2030 ridership forecast (IDOT, 2012), assuming a constant demand growth rate. Total passenger trips among the 30 OD pairs will amount to 2272 per day. Keeping the same passenger train frequency, the average schedule delay cost per passenger remains at $46.6, which is not surprising, because demand is only proportionately scaled. However, if one more passenger train is added in each direction, the average schedule delay cost will be reduced by $2.3, to $44.3/passenger.

Figure 17: Train paths on the Chicago-St Louis higher speed rail line with 2015 passenger demand scenario
The freight side traffic is also expected to significantly increase with planned infrastructure investment such as freight terminal facility development in Joliet (IDOT, 2012). To this end, we adopt IDOT’s projected freight traffic data in 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 freight side costs from solving the lower-level scheduling problem given five and six passenger trains are presented in the two right bars of Figure 18. The freight railroad will suffer significant cost increase resulting from en-route delay, departure delay, and in particular, foregone demand. Indeed, adding one passenger train results in more than doubled freight foregone demand cost. This suggests strong presence of capacity constraints on this line given passenger and freight demand growth in the future.

![Figure 18: Freight side cost on the Chicago-St. Louis higher speed rail line with different demand scenarios](image)

### 6 Concluding remarks

High performance passenger rail is gaining momentum in the United States. This paper contributes to the literature by developing a hypergraph based, strategic level planning model for mixed train operations on single-track, shared use passenger and freight corridors. The hypergraph based approach, originally introduced by Harrod (2011) to train scheduling, has proven more capable than the traditional methods to deal with path conflicts when trains transition between controlled track blocks. Given the scheduling priority of passenger trains as mandated by law in the US, a two-level modeling approach is proposed and implemented, with comprehensive and realistic consideration of different cost components. Passenger train schedules are first optimized at the upper level, where our study differs from previous research by explicitly considering and modeling passenger schedule delay, using passenger preferred departure time profiles relative to train schedules. At the lower level, we identify freight train schedules using the remaining capacity on the line. The lower-level objective is to minimize freight side costs, consisting of costs due to foregone demand, late departures, and en-route delays.

In addition to meet the conventional train operational constraints such as unique departure from origin and unique sinking at the destination for each train, block flow conservation and occupancy, resource use during transition, and minimum headway requirement, our scheduling problem further requires the order of passenger trains to be maintained on the passenger side. To this end, four approaches are proposed. The first approach penalizes violation of train order by a big number, $M$. The second approach introduces a new set of constraints that ensures the order of trains is maintained. In the third approach, we introduce new variables to represent the combination of train starting arcs, thereby transforming the original quadratic integer programming problem to a linear binary program. The fourth approach further takes advantage of the unique
structure of our problem to modify the linearized formulation. The last approach demonstrates superior performance in terms of running time and memory required to store the multiplier matrices and branch-and-cut tree.

Using realistic parameter values, our numeric experiments suggest that scheduling more passenger trains on a shared corridor lowers passenger schedule delay, at the price of freight side cost increase. The marginal freight cost increase is in most cases higher than the marginal passenger schedule delay reduction, especially when frequent passenger train services already exist on the corridor. Passenger schedule delay cost is not equally distributed among trains due to the peakedness nature of passenger preferred departure times. The heterogeneity of train speed significantly affects freight side cost, most of which comes from foregone demand. Our analysis further shows that having greater tolerance of degrading on-time performance of freight trains reduces foregone demand but leads to more delay en-route and at the departure.

The present research can be extended in several directions. First, our numerical experiment in sub-Section 5.2 shows that some passengers will be “unattended” due to train seating capacity constraints when only one train runs in each direction. How constrained train seating capacity, especially when passenger demand is high, would affect travelers’ train selection behavior, and consequently the assessment of passenger benefits, should 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) may be two separate objectives. To this end, a multi-objective approach may be more appropriate on the passenger side schedule optimization. Third, the optimization models involve simplifications. For instance, passenger trains are assumed to stop at the same stations and train acceleration and braking are ignored. Co-existence of multiple types of passenger services specified by heterogeneous train speed characterization, different stops, and possibility of overtaking could more realistically reflect train operational characteristics. The price is added modeling complexity, thus giving rise to the issue of the right tradeoff between model realism and computational efficiency. Fourth, a more sophisticated formulation encompassing the time passengers spend at the origin, destination, and in the train would further enhance the characterization of train schedule inconvenience to passengers. Fifth, the findings are based on looking into one sample problem and one prospective real-world shared use corridor. To better understand whether the findings are generalizable, more numerical experiments with varying model parameters, such as corridor distance, line configurations (location and density of sidings), and passenger PDT profiles, are warranted.

Acknowledgment

This research was partially supported by the Illinois Department of Transportation, 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 for train scheduling. We are also grateful to three anonymous reviewers for their constructive suggestions, which substantially improved the content. 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: Train paths for the small sample problem

Figure 1A: Train paths with two passenger trains

Figure 2A: Train paths with four passenger trains

Figure 3A: Train paths with six passenger trains

Figure 4A: Train paths with eight passenger trains

Figure 5A: Train paths with ten passenger trains

Figure 6A: Train paths with twelve passenger trains